From Malaria to the Sparkling of the Sea – Organelle and Host Cell Evolution in Alveolates (Apicomplexa, Dinoflagellates, Ciliates)

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina

zu Braunschweig

zur Erlangung des Grades

einer Doktorin der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

D i s s e r t a t i o n

von Ann-Kathrin Ludewig-Klingner
aus Celle

1. Referent: Privatdozent Dr. Jörn Petersen
2. Referent: Professor Dr. Michael Steinert
eingereicht am: 06.03.2017
mündliche Prüfung (Disputation) am: 04.09.2017

Druckjahr 2017

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Publikationen

Petersen, J., Ludewig, AK., Michael, V., Bunk, B., Jarek, M., Baurain, D. & Brinkmann, H. *Chromera velia*, Endosymbioses and the Rhodoplex Hypothesis – Plastid Evolution in Cryptophytes, Alveolates, Stramenopiles and Haptophytes (CASH Lineages). Genome Biology and Evolution 6 (3), 666-684 (2014).

Tagungsbeiträge

Ludewig, AK., Brinkmann H. & Petersen, J. Apicomplexans and the fate of plastids. (Vortrag) XVI. Annual Meeting of the International Society of Endocytobiology German Section (ISE-G), Herzogenhorn (2014).

Für meine Familie

„My life amounts to no more
than one drop in a limitless ocean.
Yet what is any ocean,
but a multitude of drops?"

(David Mitchell, *Cloud Atlas*)

Table of Contents

Abstract

Single-celled eukaryotes are characterized by a high degree of structural variation and evolutionary diversity. This variety is most notably represented by the group of Alveolata, which comprises photoautotrophic, heterotrophic and even parasitic species. The alveolates are composed of three different phyla: the Ciliata, the Dinoflagellata and the Apicomplexa including the malaria pathogen *Plasmodium falciparum* and the causative agent of toxoplasmosis *Toxoplasma gondii*. The basis for the conspicuous diversity within this group and other protistan lineages is their evolutionary history including a complex puzzle of successive endosymbioses. Once, symbionts became permanent residents in their hosts and the derived organelles are represented by mitochondria and plastids. Mitochondria arose from a primary endosymbiosis event with an α-proteobacterium, whereas primary plastids developed from an uptaken cyanobacterium. Complex plastids originated by secondary or even tertiary endosymbiosis by the engulfment and reduction of a eukaryotic alga. According to their bacterial origin, these cell organelles harbor nucleus-independent genomes encoding the apparatus for the maintenance of the respective compartment. Over time, most of these genes were transferred to the nucleus of the host by a process called endosymbiotic gene transfer (EGT) and particular pre-sequences mediate the re-import of the corresponding proteins into the organelle. All other organelles of the eukaryotic cell including the peroxisomes that are best known for their protection against cell damaging stress factors were established in a non-endosymbiotic context.

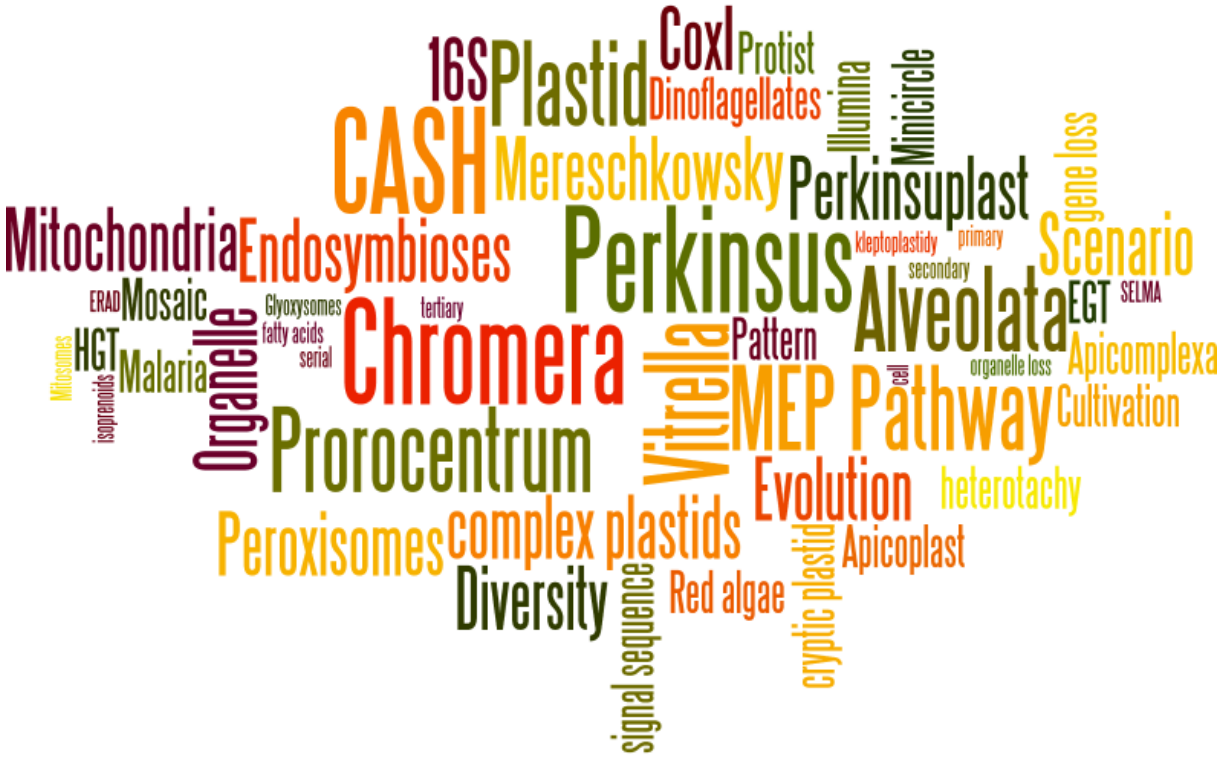
This thesis addresses the evolution of plastids and mitochondria in Alveolata, which are characterized by a strong reduction in their organelle genomes. My work based on the establishment of transcriptomes and draft genomes from three alveolate key organisms, i.e. *Chromera velia*, *Vitrella brassicaformis* and *Perkinsus olseni*, and comparative genome analyses. Hence, in the photoautotrophic alga *C. velia* (Apicomplexa) I could identify the smallest mitochondrial genome consisting of only one single protein-coding *coxI* gene. Furthermore, high-throughput genome sequencing of the oyster parasite *Perkinsus olseni* (Dinoflagellata) documented the complete loss of its plastid genome, but the maintenance of this cryptic organelle is strongly supported by the expression of nuclear-encoded proteins with characteristic bipartite targeting signals for plastid-specific metabolic pathways. Finally, this study showed the common peroxisomal ancestry of ciliates, dinoflagellates and Apicomplexa. A set of diagnostic markers provided an unequivocally proof for the presence of peroxisomes in *T. gondii*, and it moreover suggested the alga *V. brassicaformis* as a promising reference organism to study apicomplexan peroxisome biology.

Zusammenfassung

Einzellige eukaryotische Organismen zeichnen sich durch ihren hohen Grad an struktureller und evolutionärer Diversität aus. Diese Vielfalt wird vor allem durch die Gruppe der Alveolata repräsentiert, welche photoautotrophe, heterotrophe und parasitische Arten umfasst. Die Alveolata bestehen aus drei Phyla: den Ciliaten, den Dinoflagellaten und den Apicomplexa, die unter anderem den Malaria-Erreger *Plasmodium falciparum* oder auch den Toxoplasmose-Erreger *Toxoplasma gondii* umfassen. Die Grundlage für diese hohe Diversität in dieser Gruppe und in anderen Protisten bildet ihre Entstehungsgeschichte, die sich aus einem komplexen Puzzle aus aufeinanderfolgenden Endosymbiosen zusammensetzt. Die einstigen Symbionten wurden zu permanenten Bewohnern in ihren Wirten und die daraus abgeleiteten Organellen werden durch Mitochondrien und Chloroplasten vertreten. Mitochondrien entstanden dabei aus einer primären Endosymbiose mit einem α-Proteobakterium, wohingegen primäre Plastiden sich aus einem aufgenommenen Cyanobakterium entwickelten. Komplexe Plastiden entstanden durch die Aufnahme und Reduzierung eukaryotischer Algen über sekundäre oder auch tertiäre Endosymbiosen. Gemäß ihrem bakteriellen Ursprung beherbergen diese Zellorganellen eigene Genome, welche den Code für die Ausrüstung zum Erhalt des jeweiligen Kompartiments tragen. Im Laufe der Zeit wurden die meisten Gene über den sogenannten endosymbiontischen Gentransfer (EGT) in den Zellnukleus des Wirts übertragen und spezielle Präsequenzen vermitteln den Reimport der entsprechenden Proteine in das Organell. Alle anderen Organellen der eukaryotischen Zelle, wie z.B. Peroxisomen, die für ihre Schutzfunktion bei zellschädigenden Stressfaktoren bekannt sind, wurden in einem nicht-endosymbiontischen Kontext etabliert.

Diese Doktorarbeit befasst sich mit der Evolution von Plastiden und Mitochondrien in den Alveolata, die stark durch die Reduktion ihrer Organellengenome geprägt sind. Meine Arbeit basiert dabei auf der Etablierung von Transkriptomen und Genomen von drei Schlüsselorganismen der Alveolata (*Chromera velia*, *Vitrella brassicaformis* und *Perkinsus olseni*), sowie auf vergleichenden Genomanalysen. Somit konnte in der photoautotrophen Alge *C. velia* (Apicomplexa) das kleinste mitochondriale Genom bestehend aus nur einem einzigen Protein-kodierenden Gen *coxI* identifiziert werden. Durch die Hochdurchsatzsequenzierung des Genoms des Austernparasiten *P. olseni* (Dinoflagellata) konnte der vollständige Verlust seines Plastidengenoms dokumentiert werden, wohingegen der Erhalt eines kryptischen Organells durch die Expression von nukleär kodierten Proteinen unterstützt wird, welche charakteristische zweiteilige Signalsequenzen für Plastiden-spezifische Stoffwechselwege tragen. Abschließend zeigt diese Arbeit den gemeinsamen Ursprung der Peroxisomen in den Alveolata. Es werden diagnostische Marker vorgestellt, die einen eindeutigen Nachweis von Peroxisomen in *T. gondii* zeigen. Außerdem schlagen wir die Alge *V. brassicaformis* als vielversprechenden Referenzorganismus vor, der somit wesentlich zum Verständnis der Biologie von Peroxisomen in Apicomplexa beitragen könnte.

1. Introduction



An organelle is described as a well-defined reaction chamber, a compartment within a cell, which adopts specific functions. This compartmentalization of the cell allows the coexistence of multiple biochemical environments (Diekmann, Pereira-Leal 2013). Essentially, the main function of organelles is to convert energy until it is useful for the cell. For instance, the mitochondria extract energy from sugars, fats and other nutrients in order to convert them to ATP molecules with the help of oxygen (McBride et al. 2006). For good reason, mitochondria are declared as the 'powerhouses' of the cell (Siekevitz 1957). Additionally, some organisms possess plastids which can extract chemical energy from sunlight by photosynthesis. At the same time, the plastids are the location for the production and storage of essential chemical compounds (Flügge 2001). A third important organelle is represented by the peroxisome. Among the degradation of fatty acids or the biosynthesis of ether phospholipids, it detoxicates the cell from damaging metabolic products and work closely together with other cell compartments like mitochondria (Camoes et al. 2009; Islinger et al. 2010).

In the following chapters, I describe origin, evolution and functions of these three organelles with the focus on a particular protistan supergroup, the alveolates, comprising major human and livestock parasites such as the malaria agent.

Timeline of Fossil Records in Protist Research

Age	Periode	Mya	Event/Discovery	References
Cenozoic	Quarternary			
	Tertiary	15	Earliest record of *Plasmodium*	Poinar 2005
Mesozoic	Cretaceous	100	Earliest record of Apicomplexa	Poinar and Telford 2005, Poinar 2016
		114-120	Modern ciliate fossils in amber	Ascaso et al. 2005, Martin-Gonzalez et al. 2008
	Jura	190	Earliest record of diatoms	reviewed in Sims et al. 2006
		200	Modern dinoflagellate fossils	Morden and Sherwood 2002
	Trias	220-230	Modern ciliate fossils in amber	Poinar et al. 1993, Schönborn et al. 1999, Schmidt et al. 2006
		220-290	Haptophytes established	Morden and Sherwood 2002
Palaeozoic	Permian			
	Carboniferous			
	Devonian	400	Dinoflagellate-like organism	Morden and Sherwood 2002, Mogi and Kita 2010
	Silurian			
	Ordovician			
	Cambrian			
Precambrian		580	First ciliate fossils?	Li et al. 2007
		750	Main eukaryotic diversification	Parfrey et al. 2010
		1000	Primary endosymbiosis	Hoek et al.1995, Graham and Wilcox 2000, Morden and Sherwood 2002, Yoon et al. 2004
		1250	Red algae (B. pubescens)	Butterfield 2000
		1600	First ciliate fossils?	Li et al. 2009
		1800	Acritarchs (first red algae?)	Buick 2010
		2100	Acritarch: Francevillian Biota	El Albani et al. 2014
		2300	Great Oxidation Event (GOE)	Holland 2002
		3500	Kingdom Monera: Cyanobacterial fossils	Walter & Buick 1980, Altermann et al. 2006, Noffke et al. 2013
		4400	Earliest evidence of a continental crust and oceans on earth	Wilde et al. 2001

Figure 1: Historical timeline of important fossil records in protist research depicting the historical event and its reference, respectively. Inferred time periods are marked by grey dashed parentheses.

1.1 Mitochondrial Evolution

1.1.1 Primary Endosymbiosis and the Rise of Mitochondria

As the energy generation in mitochondria is an oxidative process, the origin of the mitochondrion in eukaryotes is often connected with the change of the oxygen level in the Earth's atmosphere around two billion years ago (Great Oxidation Event 'GOE': Kurland, Andersson 2000; Holland 2002; Stamati et al. 2011; Figure 1).

Mitochondria were discovered in 1856 as 'sacrosomes' by the Swiss physiologist Albert von Kölliker who observed these little particles in human muscle tissue cells. The word 'mitochondrion', which is a combination of the greek words *mitos* (thread) and *chondrus* (granule), was imposed on these organelles in 1898 by the German microbiologist Carl Benda

(van der Giezen 2011). It is commonly accepted that mitochondria arose from a primary endosymbiosis event with an α-proteobacterium (Müller, Martin 1999). Therefore, the organelle is surrounded by two membranes: the plasmamembrane of the bacterial endosymbiont and the membrane of the food vacuole generated by the phagocytotic host (Cavalier-Smith 2000; van der Giezen 2011). The high similarity of some α-proteobacterial proteins to mitochondrial proteins indicates that the uptaken α-proteobacterium was a relative of the bacterial pathogen *Rickettsia prowazekii* (Kurland, Andersson 2000; Emelyanov 2003). Different hypotheses exist of how the original host looked like: in the 'Serial Hypothesis' of Margulis (1971) the phagocytotic host already harbored a nucleus, which originated most likely from an archaebacterium, a reason why the amitochondriate Archaeozoa are the earliest branch of the eukaryotic tree (Margulis 1971; Saccone et al. 2000). In the 'Hydrogen Hypothesis' the eukaryotic nucleus and the mitochondrion were simultaneously created by the fusion of a hydrogen-requiring methanogenic archaebacterium with a hydrogen-producing α-proteobacterium (Martin, Müller 1998). The second hypothesis relies on the chimeric nature of the eukaryotic nucleus with genes of both archaebacterial and eubacterial origin (Müller, Martin 1999; Saccone et al. 2000). However, both theories imply that the endosymbiont has transferred the majority of its DNA to the nucleus via a process called endosymbiotic gene transfer (EGT). These nuclear-encoded genes contain target sequences to be transferred back to their site of action in the mitochondrion (Martin, Herrmann 1998; Kurland, Andersson 2000; Adams, Palmer 2003; Martin 2003; Timmis et al. 2004). The organelle itself retains a small DNA molecule representing the reduced bacterial chromosome. For example in humans, the mitochondrial genome is a circle of 16 kilobases (kb) harboring 37 genes: 22 tRNAs for the protein synthesis of mitochondrial encoded genes, two rRNAs (12S and 16S) and 13 genes coding for polypeptides mostly associated with the oxidative phosphorylation step, the energy generating process in mitochondria (Wallace 2005; Chan 2006).

1.1.2 Mitochondria-related Organelles (MROs)

In cells of some amitochondriate organisms, organellar structures were discovered surrounded by two membranes but lacking their own genomes (Martin, Müller 1998; Martin, Herrmann 1998; Williams, Keeling 2003; van der Giezen 2009). These structures were classified as mitochondria-related organelles and can be divided into two groups. The function of 'type I MROs' is largely unknown (Adams, Palmer 2003), but it is assumed that their sole function might be the synthesis of iron-sulfur (Fe-S) clusters which are important cofactors in diverse chemical reactions (Embley et al. 2003; Tovar et al. 2003; van der Giezen 2011). These

'cryptic mitochondria' were foremost described as 'cryptons' in the anaerobic amitochondriate parasitic species *Entamoeba histolytica* of the kingdom Amoebozoa (Mai et al. 1999). Later they were designated as 'mitosomes' in the same species based on the localization of the mitochondrial chaperonin CPN60 in these organelle-like structures (Tovar et al. 1999). On the other hand, the 'type II MROs' play a role in ATP production converting pyruvate into carbon dioxide and ATP using different enzymes than conventional mitochondria (Müller 1993; van der Giezen 2011). Along the way, they produce hydrogen as a by-product of the pyruvate oxidation that is why D. G. Lindmark and M. Müller suggested for them the term 'hydrogenosomes' in 1973 (Martin, Müller 1998; Embley et al. 2003). In contrast to the general assumption that hydrogenosomes evolved independently from mitochondria, it could be shown that both organelles are structural similar to each other and phylogenetic analyses of specific proteins showed a common ancestry (Bui et al. 1996; Dyall, Johnson 2000). Both hydrogenosomes and mitosomes are linked to the adapted lifestyle of anaerobic organisms and they are seen as the 'anaerobic variants' of conventional mitochondria (Hackstein et al. 2006). Both anaerobic variants have in common that they are involved in the synthesis of Fe-S cluster. Thus this pathway seems to be the only raison d'être in maintaining the whole organelle (Gabaldón 2012).

1.2 Plastid Evolution

1.2.1 "Per aspera ad astra": The Rise of Photosynthetic Eukaryotes

By the development and evolution of photosynthesis, sufficient energy and oxygen was provided to enable and maintain complex life on our planet. The photosynthetic life was represented first by organisms which had an anaerobic life style and utilized other electron donors like hydrogen, iron or hydrogen sulfide instead of water ('anoxygenic photosynthesis', Blankenship 2010; Hohmann-Marriott, Blankenship 2011). However, oxygenic photosynthesis was already present before the oxygen started to accumulate in the Earth's atmosphere (Buick 2008). This kind of energy generation is represented best by the robust fossil record of photosynthetic microorganisms: cyanobacterial fossils in the form of stromatolites dating back to 3.5 billion years ago (Walter et al. 1980; Des Marais 2000; Altermann et al. 2006; Blankenship 2010; Noffke et al. 2013; Figure 1).

In 1883 the German botanist A. F. W. Schimper designated the photosynthetic active organelles of land plants as 'chloroplasts' or 'plastids', and moreover he proposed that they exclusively proliferate by fission. Based on these findings, the Russian botanist Konstantin

Mereschkowsky developed in 1905 the first concept of endosymbiosis by realizing that the 'chromatophores' (plastids) are of bacterial origin. Hence, comparably to mitochondria, the origin of eukaryotic photosynthesis can be traced back to an endosymbiotic event between a bacterium, a cyanobacterium, and a non-photosynthetic eukaryotic precursor cell. The event of primary endosymbiosis was estimated around one billion years ago based on molecular clock analyses using calibration points from fossil records (Morden, Sherwood 2002; Yoon et al. 2004; Figure 1). From such an event three main lineages arose: the glaucophytes, the green algae plus land plants (Viridiplantae) and the red algae (Figure 2). Surprisingly, the time estimation of the first primary endosymbiotic event for plastids is put into another perspective by the discovery of organic-walled acritarchs which are often "[…] *interpreted as the encysted remains of eukaryotic algae"* (El Albani et al. 2014). For instance, such microfossils are assumed to be found in sediments of the 2.1 billion years old Francevillian Biota and furthermore in mid-Proterozoic rocks which are about 1.8 Ga old (Buick 2010; El Albani et al. 2014). Finally, the first multicellular fossil record of the red alga *Bangiomorpha pubescens* was dated to 1.25 Ga (Butterfield 2000). Thus, it is possible that the age of the endosymbiotic event with a cyanobacterium and the age of complex eukaryotic life is still underestimated.

In the course of primary endosymbiosis, the enslaved cyanobacterium was reduced to a cell organelle which enables the eukaryotic cell a photoautotrophic life style according to the motto 'You are what you eat' expressed by W. F. Dolittle in 1998. Primary plastids are surrounded by two membranes that are assumed to be relicts of the former endosymbiont in the host cell food vacuole (Cavalier-Smith 2000; Reyes-Prieto et al. 2007). The genome in plastids is a circular molecule representing the reduced genome of the cyanobacterial endosymbiont. For instance, the 154 kb plastid genome of the thale cress *Arabidopsis thaliana* incorporates about 87 protein-coding genes (eight of them duplicated in inverted repeat regions), four rRNAs and 37 tRNAs (Sato et al. 1999). In contrast, the red alga *Chondrus crispus* possesses a circular molecule of 180 kb with 240 protein-coding genes, three rRNA genes and 30 tRNA genes (Collen et al. 2013; Janouskovec et al. 2013). In the course of endosymbiosis, plastid genes were either transferred to the nucleus by EGT or replaced by nuclear homologs. Similar to mitochondria, nuclear-encoded plastid proteins incorporate target sequences for their return transport to the organelle. Consequently, the plastome retains genes which are not yet established in the nucleus or could not be transferred due to functional barriers (Cavalier-Smith 2000).

The third lineage containing primary plastids are the glaucophytes with its most prominent representative *Cyanophora paradoxa*. Their plastids differ essentially from other primary

plastids by containing a peptidoglycan layer, a relic of the uptaken cyanobacterium and a reason why they carry the name 'cyanelle' (Stirewalt et al. 1995). In addition to the conventional chloroplast gene set, these organelles encode some of the phycobilisome structural genes which are light harvesting antennae of the photosystem II and which are also found in cyanobacteria and red algae (Glazer 1985; Lambert et al. 1985).
However, although there are differences between the primary plastids of glaucophytes and the red algae/Viridiplantae, their origin is assumed to be monophyletic which is supported by numerous multigene analyses (Rodriguez-Ezpeleta et al. 2005; Hackett et al. 2007; Burki et al. 2008; Burki et al. 2009) and they are designated as 'Archaeplastida' (Adl et al. 2005). Hence, it is commonly accepted that primary plastids originated from one single uptake of a cyanobacterium.

1.2.2 Endosymbiotic Origin of Complex Plastids

The process of endosymbiosis was the key mechanism in protozoan plastid evolution. However, within unicellular eukaryotic organisms, plastids can be structural dissimilar. In contrast to mitochondria where a single endosymbiotic uptake of a bacterium give rise to all known mitochondria (Gray et al. 2001), some plastids are more complex than others and their origin can be construed from serial endosymbiosis which is reminiscent of the nesting principle of the Russian Matryoshka dolls (Petersen et al. 2006). Thus, complex plastids are surrounded by three to four membranes referring to secondary, tertiary or even quarternary endosymbiosis. Morden and Sherwood asked rightly in 2002: "[…] *how many endosymbiotic events are required to account for all diversity present among photosynthetic algae?*" Unfortunately, the amount of endosymbiotic events is difficult to assess.
Secondary endosymbioses are demonstrated best within two protozoan lineages, which still harbor the nucleus of the engulfed endosymbiont in addition to their own nucleus, the so-called 'nucleomorph'. Hence, the lineage of Chlorarachniophyta (Rhizaria) that comprises the model organism *Bigelowiella natans* arose by a secondary endosymbiosis with a green alga, whereas the Cryptophyta (e.g. *Guillardia theta*) developed from a secondary endosymbiosis event with a red alga (reviewed in: Delwiche 1999; Douglas, Penny 1999; Cavalier-Smith 2000; Douglas et al. 2001; Gilson et al. 2006; Archibald, Lane 2009; Kim, Archibald 2008; Figure 2).

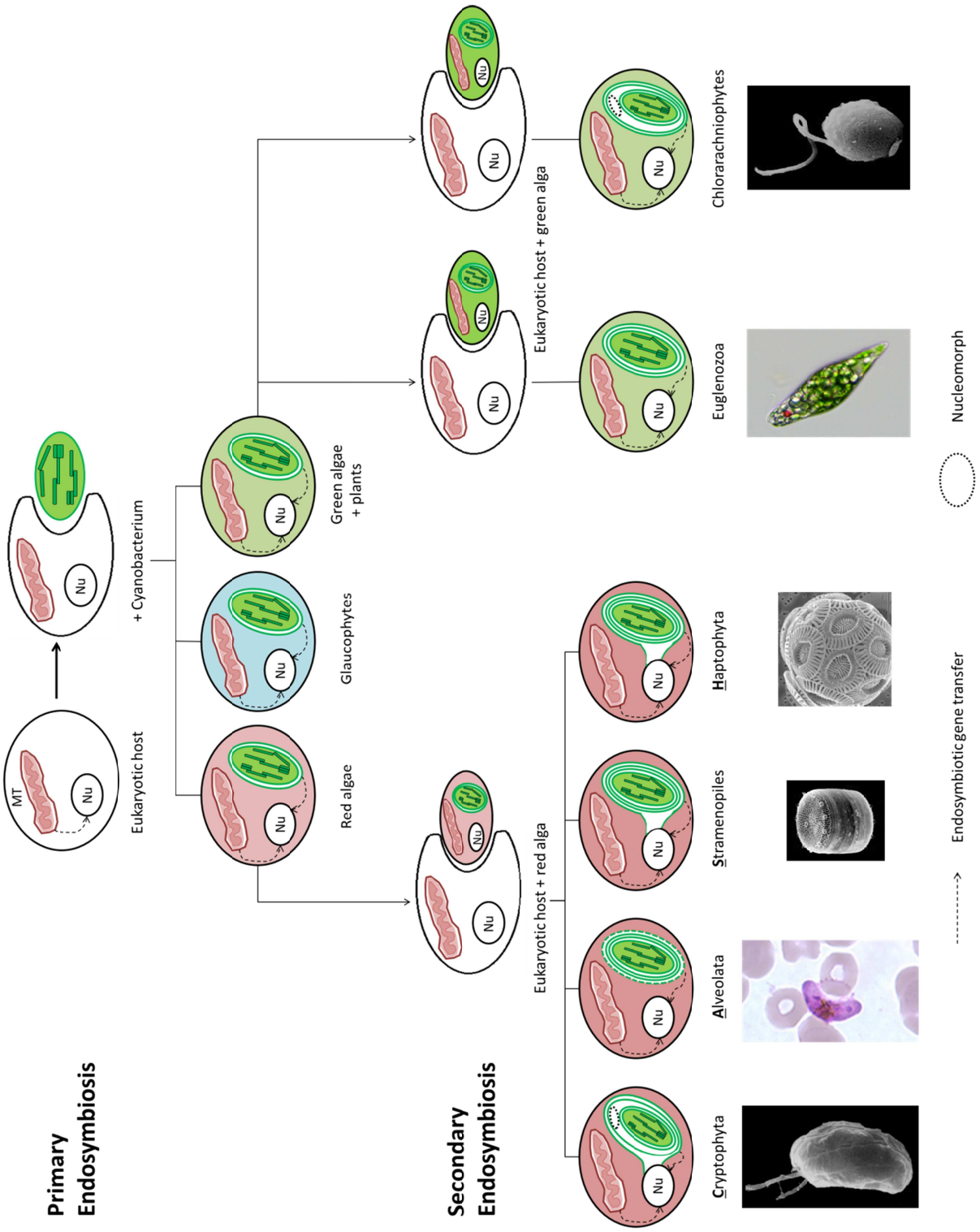

Primary Endosymbiosis
MT
Nu
Eukaryotic host
+ Cyanobacterium
Red algae
Glaucophytes
Green algae + plants
Secondary Endosymbiosis
Eukaryotic host + red alga
Eukaryotic host + green alga
Cryptophyta
Alveolata
Stramenopiles
Haptophyta
Euglenozoa
Chlorarachniophytes
Endosymbiotic gene transfer
Nucleomorph

Figure 2: Schematic illustration of the primary and secondary endosymbioses with the consequential emerged lineages. Complex algae emanating from an endosymbiotic event with a red alga are portrayed by red cells, whereas lineages originated by an endosymbiotic event with a green alga are illustrated by green cells. The single-celled glaucophytes are assigned to the blue color. The chloroplasts/complex plastids are highlighted by the strong green color; mitochondria by a light red color. Dashed arrows symbolize the genetic transfer between the endosymbiont and the nucleus of the organism by endosymbiotic gene transfer (EGT). Dashed circles are subject to the retained nuclei of the endosymbionts, the so-called 'nucleomorphs'. Images courtesy: Cryptophyta (electron scanning micrograph of *Guillardia theta*, Dr. David Hill), Alveolata (Gametocyte of *Plasmodium falciparum* in thin blood smear, CDC – Centers for Disease Control and Prevention; https://www.cdc.gov/dpdx/malaria/gallery.html), Stramenopiles (electron scanning micrograph of the diatom *Thalassiosira pseudonana*, Dr. Nils Kröger), Haptophyta (Coccoliths in an electron scanning micrograph of *Emiliania huxleyi*, Dr. Alison R. Taylor), Euglenozoa (LM-micrograph of *Euglena gracilis*, Dr. Ralf Wagner), Chlorarachniophytes (electron scanning micrograph of *Bigelowiella natans*, Dr. David Hill). Abbreviations: MT = Mitochondrion, Nu = Nucleus.

From a phylogenetic point of view, it is assumed that a secondary endosymbiosis event with a green alga took place twice independently represented by the lineages chlorarachniophytes and Euglenozoa (Cavalier-Smith 1999; Rogers et al. 2007; Turmel et al. 2009; Figure 2). In comparison to the chlorarachniophytes with four plastid membranes, the plastid of Euglenozoa (e.g. *Eulgena gracilis*) is surrounded by only three membranes and they do not harbor a plastid-associated nucleomorph (Leander et al. 2001).

In addition to the Cryptophyta, three further lineages evolved from a secondary endosymbiosis with a red alga: the Haptophyta (e.g. *Emiliania huxleyi* [synonym: coccolithophores]; Figure 2), the stramenopiles (e.g. diatoms; Figure 2) and the Alveolata (e.g. the malaria agent *Plasmodium falciparum*, Figure 2). Their plastids are surrounded by three (dinoflagellates, Alveolata) or four membranes and in contrast to the Cryptophyta they lack a nucleomorph. Furthermore, the outer membrane (phagosomal membrane) in Cryptophyta, Haptophyta and stramenopiles is continuous with the nuclear envelope placing the plastid within the rough endoplasmatic reticulum (Figure 2; Cavalier-Smith 2000). The phylogenetic relationship of complex plastids of red algal origin still remains cryptic and is controversially discussed (Cavalier-Smith 1999; Bodył 2005; Petersen et al. 2006; Baurain et al. 2010; Petersen et al. 2014). One central discussed hypothesis relies on the assumption that an endosymbiosis is a very rare event. Thus, in the case of red-algal derived plastids, Cavalier-Smith inferred that it would be most parsimonious if all complex 'red' plastids arose from a single secondary endosymbiosis with a red alga. On this account, he summarized the group of so-called 'chromists' (haptophytes, cryptophytes and stramenopiles) and the group of alveolates (ciliates, Apicomplexa and dinoflagellates) to the presumable superensemble of 'chromalveolates'. Based on the common characteristic that chlorophyll c is present in all chromalveolate lineages, he postulated in his 'Chromalveolate Hypothesis' that these lineages are monophyletic implicating that all non-photosynthetic aplastidic lineages within the chromalveolates (e.g. ciliates of the alveolates and oomycetes within the stramenopiles) have lost their plastids secondarily (Cavalier-Smith 1999). In several studies, the Chromalveolate

Hypothesis was strongly supported by phylogenetic analyses of plastid-encoded and nuclear-encoded plastid-targeted genes (Fast et al. 2001; Yoon et al. 2002; Harper, Keeling 2003; Patron et al. 2004; Bachvaroff et al. 2005; Petersen et al. 2006), whereupon analyses of nuclear markers representing the host cell showed only weak support, because haptophytes and cryptophytes cannot be robustly located and often branch in different regions of the tree (Ali et al. 2001; Harper et al. 2005; Hackett et al. 2007; Patron et al. 2007; Burki et al. 2008; Burki et al. 2009). Despite the discrepancy between plastid and nuclear phylogenies, the term chromalveolates was commonly used in publications and even introduced into the microbial taxonomy (Adl et al. 2005). However, in 2010, the Chromalveolate Hypothesis was falsified in a large phylogenomic study using new techniques of data analysis (Baurain et al. 2010) and the chromalveolates were subsequently removed from taxonomic classification (Adl et al. 2012). Based on the incongruences of plastid and nuclear phylogenies, Baurain et al. suggested to replace the misleading name chromalveolates by the descriptive term 'CASH' (Cryptophytes, Alveolata, Stramenopiles and Haptophytes) lineages (Petersen et al. 2014; Figure 2). Multigene analyses showed that complex algae with red plastids are polyphyletic and these analyses moreover supported a new monophyletic superensemble of protistan phyla: Stramenopiles, Alveolata and Rhizaria (SAR, Burki et al. 2008; Burki et al. 2009; Hampl et al. 2009; Burki et al. 2010; Parfrey et al. 2010). In 2014, we proposed the 'Rhodoplex Hypothesis' to describe the plastid evolution in the CASH lineages. Our concept is consistent with the CASH nomenclature and the independent host cell origin of all four lineages. Furthermore, we postulated that all complex plastids of red algal origin arose from a single secondary endosymbiotic event, but with successive higher order eukaryote-to-eukaryote endosymbioses (EEE, Baurain et al. 2010) in order to describe the incongruence between plastid and host cell markers (Petersen et al. 2014).

The nuclei of protists with complex plastids have a mosaic composition because of (1) genes obtained by non-endosymbiotic horizontal gene transfer (HGT), which can replace plastid-encoded genes (e.g. in dinoflagellates; Nosenko et al. 2006) and (2) because they have recruited many genes from their uptaken endosymbionts by EGT (Timmis et al. 2004). In the case of assumed secondary complex plastids, these nuclear-encoded plastid genes usually contain bipartite signal sequences for the (re)transport of the corresponding proteins into the organelle. The signal sequences consist of a signal peptide that ensures the passage through the outer membranes of the complex plastid whereas the following transit peptide navigates the protein through the residual membranes of the plastid after the signal peptide is cut off

(Douglas 1998; Gray 1999; Cavalier-Smith 1999; Archibald et al. 2003; Harper, Keeling 2003; Burki et al. 2012; Chan et al. 2012).

Tertiary endosymbioses of a eukaryotic cell with stramenopile (Chesnick et al. 1996; Chesnick et al. 1997; Inagaki et al. 2000; Saldarriaga et al. 2001; Burki et al. 2014), cryptophyte (Wilcox, Wedemayer 1985; Hackett et al. 2003) and haptophyte (Tengs et al. 2000) endosymbionts were already demonstrated in the highly diverse phylum Dinoflagellata within the alveolate supergroup (Yoon et al. 2005).

1.3 Peroxisomal Evolution

Peroxisomes, also known as 'microbodies', are organelles surrounded by a single membrane which are primarily required for the breakdown of long-chain fatty acids (Rhodin 1954; Baudhuin et al. 1965; de Duve, Baudhuin 1966; Gabaldon 2010). Furthermore, this compartment protects the cell of oxidative damage assembling metabolic pathways which generate continuously high reactive hydrogen peroxide (H_2O_2) that can be rapidly deactivated by peroxisome-localized oxidoreductases such as the catalase (de Duve, Baudhuin 1966; Reddy, Mannaerts 1994). Peroxisomes with specific cellular functions and metabolic capacities are known as 'Woronin bodies' in filamentous fungi (Markham, Collinge 1987), 'glycosomes' in trypanosomes (Opperdoes, Borst 1977) and 'glyoxysomes' in fungi, land plants and some protists (Kornberg, Krebs 1957; McCammon et al. 1990). The composition of metabolic pathways within peroxisomes can vary, even in different cellular stages and tissue types, which is provoked by the adaptation of this organelle to the specific needs of the cell (Hayashi et al. 2000; Smith, Aitchison 2009; Gabaldon 2010). However, the 'peroxins' represent a constant protein pool for the maintenance of peroxisomes. They are responsible for the peroxisome biogenesis, protein import, translocation and recycling as well as peroxisome proliferation (Smith, Aitchison 2013). Mutations in the peroxins were shown to cause fatal neurological disorders in humans (Schlüter et al. 2006; Smith, Aitchison 2013). Nuclear-encoded peroxisomal matrix proteins incorporate peroxisomal targeting sequences (PTS) which mediate their transport to the organelle (Rucktäschel et al. 2011). The conserved C-terminal tripeptide PTS1 has the consensus motif (S/A/C)-(K/R/H)-(L/M) and is the most observed import signal (Lametschwandtner 1998). In contrast the N-terminal nonapeptide PTS2 is slightly difficult to determine due to its low conservation status. Anyway, the consensus motif [(R/K)-(L/V/I)-X_5-(H/Q)-(L/A)] was defined with 'X_5' as five discretionary amino acids (Rachubinski, Subramani 1995).

The origin of peroxisomes has been controversially discussed and can be summarized by two possible scenarios. Peroxisomes proliferate comparably to mitochondria and plastids by fission and the first scenario thus proposes an enslavement of a prokaryotic cell by endosymbiosis. Unlike mitochondria and plastids, peroxisomes do not contain their own genome and the single surrounding membrane would also argue against an endosymbiotic scenario. However, Lazarow and Fujiki explained this phenomenon by the complete transfer of peroxisomal genetic material to the nucleus and the loss of additional membranes (Lazarow, Fujiki 1985). With regard to an endosymbiotic scenario actinobacteria were proposed as possible endosymbiotic donors, but a subsequent study disproved this assumption and showed that the conclusion was drawn based on phylogenetic artifacts (Duhita et al. 2010; Gabaldon, Capella-Gutierrez 2010).

The second scenario that proposes a genuine eukaryotic invention of peroxisomes is meanwhile largely preferred compared to the endosymbiotic hypothesis in contemporary review articles (Gabaldón 2014; Speijer 2015). Some peroxins of the peroxisomal protein import system are homologous to proteins of the endoplasmatic-reticulum-associated protein degradation (ERAD) system in the endoplasmatic reticulum (Schlüter et al. 2006; Gabaldón et al. 2006; Bolte et al. 2011). Furthermore, it could be shown that mutations in peroxins resulted in peroxisome-deficient strains of the fungal model organism *Saccharomyces cerevisiae* (Erdmann et al. 1989; Erdmann, Kunau 1992; van der Leij et al. 1992). However, peroxisome-loss is reversible and complementation with the wild type genes resulted in the *de novo* formation of the organelle by offshoots of the ER (Hoepfner et al. 2005; Fagarasanu et al. 2007).

The functional role of peroxisomes has been intensively studied in land plants, yeast and mammals (Cross et al. 2016; Veenhuis, van der Klei 2014; Fujiki et al. 2014) and it was also investigated in heterotrophic organisms like trypanosomes (Euglenozoa), oomycetes (early branching group of the stramenopiles) and ciliates (early branching group of Alveolata) (de Duve, Baudhuin 1966; Philippi et al. 1975; Opperdoes, Borst 1977; Adl et al. 2012). In autotrophic protists peroxisomes were studied in the diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* (Armbrust et al. 2004; Gonzalez et al. 2011). However, the knowledge about the distribution of peroxisomes among protists is sparse even if they were supposed to be present in nearly all eukaryotic lineages (Gabaldon 2010).

1.4 Alveolata

Alveolata are unicellular eukaryotic microorganisms including ciliates (e.g. *Paramecium*), the parasitic group of Apicomplexa (e.g. the malaria pathogen *Plasmodium*) and the dinoflagellates (main component of the phytoplankton). The name of this group can be traced back to cortical alveoli, flattened vesicles in close vicinity to the plasmamembrane of the organisms which can be got lost secondarily in some species. Other characteristics are micropores and the tubular cristae of the mitochondria (Adl et al. 2012).

Alveolates form a well-supported 'superensemble' and monophyletic group based on phylogenetic analyses (Morse et al. 1995; Janouskovec et al. 2010; Bachvaroff et al. 2014). Thereby the ciliates form the earliest-branching alveolate phylum to the sister-group of Apicomplexa and dinoflagellates (Gajadhar et al. 1991; Cavalier-Smith 1993; Fast et al. 2002; Leander, Keeling 2004; Adl et al. 2012).

1.4.1 Ciliata

Species Diversity

The word 'ciliate' is related to the human eyelash and indicates one of the most evident characteristics of this alveolate phylum. The surface of ciliates is in general laced with little hair-like structures called 'cilia' which serve as a form of locomotion and food ingestion (Prescott 1994; Gao, Katz 2014). Ciliata are heterotrophic ingesting a rash of organisms such as bacteria and algae. Via 'filter-feeding' the food particles are concentrated by a position-specific accumulation of cilia around the oral groove, the 'mouth' of ciliates. Afterwards endocytosis takes place and the food is absorbed in food vacuoles (Hausmann 2002).

Another characteristic of ciliates is the nuclear duality within the cells. They possess (1) one to several diploid germline nuclei ('micronucleus') for sexual reproduction via conjugation (exchange of genetic material between two compatible mating types) and (2) one to several polyploid somatic nuclei ('macronucleus') which are formed by micronuclei and are involved in the asexual reproduction of the cell by fission (Prescott 1994).

The phylum Ciliata is a very large and diverse group incorporating twelve (Adl et al. 2012) to fourteen (Gao et al. 2016) different classes with approximately 3,500 extant known species (Adl et al. 2007). Two main branches are defined within this phylum: the subphylum Postciliodesmatophora and the subphylum Intramacronucleata (Gao, Katz 2014). The subphylum Postciliodesmatophora comprises two classes, the Karyorelictea (macronuclei do not divide, only the micronuclei) and the Heterotrichea (macronuclei divided by extra-macronuclear microtubuli, e.g. the genus *Condylostoma*) (Adl et al. 2012; Gao et al. 2016),

which were documented as monophyletic groups based on phylogenetic analyses (Gao, Katz 2014; Chen et al. 2015). The subphylum Intramacronucleata can be divided into two main groups: (1) the 'CONthreeP' incorporating the six classes Colpodea, Oligohymenophorea (e.g. the well-known genera *Tetrahymena*, *Ichthyophthirius*, *Paramecium*), Nassophora, Phyllopharyngea (e.g. *Chilodonella*), Plagiopylea and Prostomatea; and (2) the 'SAL' group with the three classes Spirotrichea (e.g. the genera *Oxytricha*, *Stylonychia* and *Favella*), Armophorea and Litostomatea. Unfortunately these two groups are only weakly supported by morphological synapomorphies (Gao et al. 2016). However, the Intramacronucleata are characterized by the eponymous trait that the macronuclei are divided by intramacronuclear microtubuli (Adl et al. 2012).

Recently, a new class was added: the Mesodiniea incorporating the former phylogenetically problematic and not classified genus *Mesodinium* (Chen et al. 2015; Gao et al. 2016). Although, this class branched strictly as the sister to all ciliates in rRNA-trees (Gao et al. 2016), the monophyly within the genus *Mesodinium* could not be confirmed on a phylogenomic level (Chen et al. 2015).

Furthermore, monophyletic groups were confirmed by rRNA-phylogenies for the ciliate classes Phyllopharyngea, Karyorelictea, Armophorea, Protomatea, Plagiopylea, Colpodea and Heterotrichea (Gao et al. 2016).

Fossil Records

Ciliate fossils are rare because they lack hard cellular structures which could otherwise easily fossilize. However, the marine group of planktonic tintinnids (class: Spirotrichea) is the sole ciliate lineage which has a significant historical record because of their fossilized loricae (protective outer covering; Dunthorn et al. 2015). Frequently records of soft-bodied ciliate fossils in Triassic and Lower Cretaceous amber are documented (Poinar et al. 1993; Schönborn et al. 1999; Ascaso et al. 2005; Schmidt et al. 2006; Martin-Gonzalez et al. 2008; Dunthorn et al. 2015; Figure 1).

The supposedly oldest fossils of ciliates were described from Proterozoic rocks of the Doushantuo Formation in China dating back to 580 mya (Li et al. 2007; Figure 1). Later on, it was also speculated that the oldest tintinnid records were found in Mesoproterozoic rocks from Central China dating even back to 1.6 Ga (Li et al. 2009; Figure 1). However, these findings were disputed and reassessed as *incertae sedis* eukaryotes or even inorganic particles (Dunthorn et al. 2010; Dunthorn et al. 2015).

1.4.2 Dinoflagellata

Species Diversity

Dinoflagellates are important photosynthetic symbionts of corals and contribute to the marine productivity (Lin 2011; Wisecaver, Hackett 2011). This alveolate phylum is highly diverse and can be classified into two major lineages: the Syndiniales and the so-called core dinoflagellates or dinokaryotes (Hoppenrath, Leander 2010; Okamoto et al. 2012; Bachvaroff et al. 2014). The non-photosynthetic order Syndiniales, comprising *inter alia* the genera *Amoebophrya* and *Syndinium*, differ from the typical nucleus of other dinoflagellates in their low number of non-fibrillar V-shaped chromosomes (Gomez et al. 2010; Adl et al. 2012). In contrast dinokaryotes have condensed 'arched fibrillar' chromosomes with a high degree of gene duplication (Bachvaroff, Place 2008; Bachvaroff et al. 2009; Shoguchi et al. 2013; Bachvaroff et al. 2014). The core dinoflagellates incorporate the Noctilucales (e.g. *Noctiluca*) and the dinophyceaen orders Gymmnodiniales (e.g. *Lepidodinium, Nematodinium, Amphidinium, Karenia, Karlodinium, Sclerodinium, Protodinium* and *Symbiodinium*), Peridiniales (e.g. *Alexandrium, Gonyaulax, Lingulodinium, Pyrocystis, Blastodinium, Peridnium, Pfiesteria, Scrippsiella*), Dinophysiales (e.g. *Dinophysis*) and Prorocentrales (e.g. *Prorocentrum*) (Adl et al. 2012). In ribosomal phylogenies *Amphidinium carterae* was positioned as the earliest branch of all dinokaryotes and monophyletic groups were described for the Suessialeans (e.g. *Symbiodinium*), Kareniaceae (e.g. *Karenia* and *Karlodinium*), *Prorocentrum* and the Gonyaulacoids (e.g. *Alexandrium, Lingulodinium, Pyrodinium*) (Bachvaroff et al. 2014).

In some dinokaryotes the cell is covered by an amphiesma or cortex, which means that the alveoli are filled with cellulose to solid flattened thecal plates which are also known as the 'armor' of the cell (Loeblich, Loeblich 1985; Orr et al. 2012). The position of the athecate dinoflagellate species *A. carterae* as the sister to all other dinokaryotes suggests that the thecate dinoflagellates arose from an athecate ancestor (Orr et al. 2012; Bachvaroff et al. 2014).

Numerous dinoflagellates are heterotrophic and adapted to micropredation and parasitism (Coats 1999; Gornik et al. 2015). The non-photosynthetic parasitic dinoflagellate species *Oxyrrhis marina* and the genus *Perkinsus* diverge from the base of all other dinoflagellate lineages (Siddall et al. 1997; Saldarriaga et al. 2003; Bachvaroff et al. 2011). Recent phylogenetic analyses put species of the genus *Perkinsus* at the outermost position to all other dinoflagellate species, followed by *O. marina* nesting the sister group of Syndineans and the core dinoflagellates (Bachvaroff et al. 2014). However, the general acceptance is that the

common ancestor of all dinoflagellates was once photosynthetic and that the non-photosynthetic lineages lost photosynthesis or the whole plastid secondarily and independently (Saldarriaga et al. 2001; Janouskovec et al. 2010; Bachvaroff et al. 2014; Gornik et al. 2015).

Fossil Records

The first fossil record of dinoflagellate-like organisms in upper Silurian sedimentary rocks of Tunesia dated to 400 mya (Figure 1) is disputed. The pattern of armored plates (tabulation) and the archeopyle (opening on the wall) of the found cysts in the so named species *Arpylorus antiquus* were reminiscent of dinoflagellate characters and were assigned to early forms of this alveolate phylum (Calandra 1964; Sarjeant 1978). This record was generally adopted in the literature (e.g. Morden, Sherwood 2002; Mogi, Kita 2010), but it was challenged in 2012 by a French working group. The respective sediments were reanalyzed using new material and modern biogeochemical methods. With respect to these new techniques, the former dinoflagellate tabulation structures were proposed to be rather a storage structure produced by invertebrates (Le Herisse et al. 2012).

Fossils of modern dinoflagellates in form of dinocysts are mainly found in Mesozoic sediments of the late Triassic (~200 mya) after the Haptophyta were established (220-290 mya, Figure 1, Fensome et al. 1994; Morden, Sherwood 2002) and before the diatoms were arising (190 mya, Figure 1, Sims et al. 2006).

1.4.3 Apicomplexa

Species Diversity

The phylum Apicomplexa comprises about 6000 obligate parasitic species, among them the most known representatives the malaria agent *Plasmodium falciparum* and the causative organism for toxoplasmosis *Toxoplasma gondii* (Ginger 2006; Adl et al. 2007; Lim, McFadden 2010; Adl et al. 2012). Both species represent the main classes of this phylum: the Aconoidasida and the Conoidasida, respectively. The Apicomplexa are characterized by their eponymous apical complex, an organ located at the anterior end of the parasitic cell. It includes a polar ring, rhoptries, micronemes as well as sub-pellicular microtubules, and it is supposed to play a role during the invasion of an apicomplexan parasite into its host cell (Levine 1971; Adl et al. 2012; Katris et al. 2014).

In the Conoidasida an additional apical motile organelle, the conoid, is described which actively contributes to the motility and invasion of the parasite (Graindorge et al. 2016).

Coccidian parasites are the largest group of the Conoidasida incorporating *inter alia* the genera *Cyclospora*, *Eimeria*, *Hammondia*, *Neospora*, *Sarcocystis* and the most prominent *Toxoplasma,* infecting vertebrate hosts (Adl et al. 2012). The subclass Gregarinasina including amongst others the genera *Ascogregarina*, *Blabericola*, *Gregarina* and *Lankesteria*, is a known parasitic group infecting invertebrates and early diverging in the apicomplexan phylum (Leander et al. 2003). They share this early positioning with the genus *Cryptosporidium*, the agent of the cryptosporidiosis in Metazoa (Carreno et al. 1999; Barta, Thompson 2006).

The Aconoidasida are lacking a conoid in all but one cell stage, the motile gliding ookinetes. This apicomplexan class comprises two large parasitic groups: the subclass Piroplasmida including the genera *Babesia* and *Theileria,* as well as the subclass Haemosporidia which comprises the genera *Haemoproteus*, *Leucocytozoon*, *Mesnilium* and the most prominent *Plasmodium* (Adl et al. 2012).

In phylogenomic approaches Piroplasmida and Haemosporida are sisters to each other with the coccidian parasites as the next closely related group (Kuo et al. 2008). The relationship between Gregarinasina and *Cryptosporidium* as the sister to all apicomplexan parasites is still under debate. Phylogenies of small subunit rRNAs and the heat shock protein Hsp90 revealed a sister group relationship of both early diverging groups (Carreno et al. 1999; Leander, Keeling 2004), whereas other studies of the small subunit rRNAs or the actin protein show the genus *Cryptosporidium* as the earliest branching apicomplexan lineage followed by the close related gregarines (Leander et al. 2003; Leander, Keeling 2004). Contradictory results of single gene phylogenies usually arise from incomplete taxon sampling or a limited resolution of the marker genes.

The discovery of two photosynthetic apicomplexan representatives, the two algae *Chromera velia* and *Vitrella brassicaformis*, made progress in the question of a possible photosynthetic ancestry of apicomplexan parasites as well as in the question of a common photosynthetic ancestry of alveolates classifying them as important ‘connecting links’ in alveolate evolution. Both so called ‘chromerids’ were isolated from reef and coral sites of the Australian coast (Moore et al. 2008; Obornik et al. 2012).

Phylogenetic analyses placed the chromerids as the earliest branching apicomplexan lineage either in a serial branching order or as a sister group to the Apicomplexa (Janouskovec et al. 2010; Woo et al. 2015).

Historical and Fossil Records

The etiopathology of Malaria in humans was already observed and described by the 'father of modern medicine', Hippocrates, in the 5th century BC (Pappas et al. 2008). The parasite itself was described for the first time in 1880 by Alphonse Laveran in blood smear of a diseased soldier (Cox 2002).

First apicomplexan fossil records are based on mosquitos encased in amber (Poinar, Telford 2005; Poinar 2005a, 2005b; Poinar 2012; Poinar 2016). In 2005, today's extinct apicomplexan species *Paleohaemoproteus burmacis*, a relative to the extant genus *Haemoproteus* (Apicomplexa, Haemosporidia), was discovered in the abdominal cavity of a female biting midge of the genus *Protoculicoides* in a piece of amber from Myanmar dating back to 100 mya (Poinar, Telford 2005; Poinar 2016; Figure 1). Inside the culicine mosquito *Culex malariager* from a mid-Tertiary (15 mya) Dominican amber, George Poinar found the earliest record of the genus *Plasmodium* by detecting various stages of *Plasmodium dominicana* (Poinar 2005a, 2005b; Figure 1). It is assumed that the original vertebrate hosts of the progenitors of the current malaria pathogens were reptiles of the order Squamata before the change to mammal hosts (Yotoko, Elisei 2006; Poinar 2016). Because the age of these parasites can be traced back to the age of the dinosaurs, in a public interview George Poinar supposed that these gigantic reptiles were already plagued with the malaria agent and that the parasites were possibly jointly responsible for the decline of dinosaur populations (Poinar 2016; Whiteman 2016).

1.4.4 Highly Derived Mitochondrial Genomes in Alveolata

Mitochondrial genomes can vary in size as well as in gene arrangement. Alveolates usually possess a very small and fragmented organelle genome with the most reduced molecules in the phyla Apicomplexa and Dinoflagellata. Alongside the adaptation of some alveolate organisms to a parasitic life style numerous mitochondrial pathways are modified or lost due to the accessibility of host-cell metabolites replacing important steps or reaction products of these pathways (van Dooren et al. 2006; Seeber et al. 2008; Polonais, Soldati-Favre 2010; Danne et al. 2013).

Ciliata

The ciliates possess a more or less conventional mitochondrial genome: gene-rich and relatively large (Swart et al. 2012). The linear mitochondrial genomes for the oligohymenophoran species *T. pyriformis* (Burger et al. 2000), *T. tetrahymena* (Brunk 2003)

and *P. tetraurelia* (Pritchard et al. 1990) as well as for the spirotrichous species *Euplotes* sp. (Graaf et al. 2009) were intensively studied. They are of around 41-48 kb in size encoding for more than 50 genes: on average 21 protein-coding genes, 22 ciliate-specific open reading frames (ORFs), the two rRNAs for the large and small subunit (LSU and SSU) as well as seven tRNAs. Thereby, all protein-coding genes represent enzymes of the oxidation phosphorylation pathway (Pritchard et al. 1990; Burger et al. 2000; Brunk 2003; Graaf et al. 2009). Ciliate mitochondrial genomes are rapidly evolving genomes and they exhibit highly divergent protein-coding genes that could often not be assigned to a known function (Burger et al. 2000; Gray et al. 2004; Moradian et al. 2007).

Recently, the mitochondrial genome of another spirotrichous ciliate was described. In *Oxytricha trifallax*, the largest known mitochondrial chromosome (~70 kb, Genbank JN383843) with an additional linear mitochondrial plasmid (~ 5 kb, Genbank JN383842) was discovered which share the same type of telomeric repeats. The large mitochondrial chromosome incorporates 29 protein-coding genes, eleven tRNAs and two rRNAs (Swart et al. 2012).

Anaerobic ciliates usually contain hydrogenosomes instead of conventional mitochondria. The first example of a hydrogenosomal genome of mitochondrial descent was reported for the heterotrichous cockroach ciliate *Nyctotherus ovalis* (Akhmanova et al. 1998).

Dinoflagellata

The mitochondrial genomes of the phylum Dinoflagellata are as diverse as their species variety. In most of the species the genomes are expanded and rearranged by gene duplications and gene recombinations (Waller, Jackson 2009). They are distended by a high content of non-coding and repetitive DNA and highly fragmented (Nash et al. 2007; Wisecaver, Hackett 2011; Figure 3). On the other hand they contain a strong reduced gene content composed of the three protein-coding genes *cytochrome c oxidase subunit I* (*coxI*), *cytochrome c oxidase subunit III* (*cox3*) and *cytochrome b* (*cob*) of the oxidative phosphorylation pathway (Figure 3) as well as the two fragmented SSU and LSU rRNAs. Their mitochondrial genomes have in general lost the multiprotein complex I NADH dehydrogenase of the electron transport chain which is a shared characteristic with the Apicomplexa. Accordingly, this loss likely occurred in a common ancestor of Dinoflagellates and Apicomplexa (Gardner et al. 2002; Waller, Jackson 2009). The expression of mitochondrial genes is more complex than in other protists due to RNA editing, trans-splicing and non-canonical start- and stop-codons (Gray 2003; Jackson et al. 2007; Nash et al. 2008). A new mechanism of mitochondrial gene

expression with a unique genetic code was discovered in the earliest branching dinoflagellate genus *Perkinsus*. In this case the gene transcripts of *coxI* and *cob* possess extensive frameshifts at every AGG (Glycine) and CCC (Proline) codon. Two possible mechanisms for the translation of these frameshifts were proposed: (1) ribosomes skipping the first bases of these codons or (2) specialized tRNAs recognizing for example the quadruplet AGGY and the quintuplet CCCCU (Masuda et al. 2010). Deviations of these non-triplet codons within each *Perkinsus* species were observed (Masuda et al. 2010; Zhang et al. 2011). This trait prompted Zhang et al. to classify the perkinsids as an independent sister lineage ('Perkinsozoa') to all other dinoflagellates. However, the mitochondrial genome analysis of the deep branching dinoflagellate parasite *Hematodinium* sp. of the order Syndiniales which infects the hemolymphs of crustaceans showed "[…] *that much of the radical reorganization of this organelle genome occurred early in dinoflagellate radiation.*" (Jackson et al. 2012).
The smallest mitochondrial genome that was known until 2014 is found in another early-branching dinoflagellate species: *Oxyrrhis marina*. It only harbors two genes of which one gene is a fusion of the *cox3*- and the *cob*-reading frame (Slamovits et al. 2007; Waller, Jackson 2009; Figure 3).

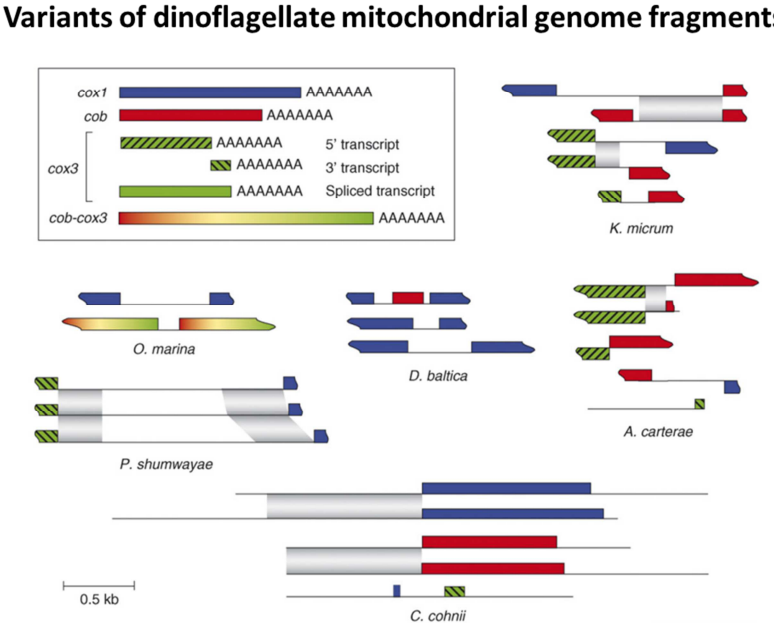


Figure 3: Schematic diagram according to Nash et al. 2008 showing genetic arrangements of mitochondrial transcripts in six different dinoflagellate species. Only fragments for the protein coding *coxI*, *cox3* and *cob* genes are shown, ribosomal sequences were discarded. The color codes for the genes are described in the box. The gray color indicates homologous non-coding regions. Genera abbreviations: *A. carterae* = *Amphidinium carterae*; *C. cohnii* = *Crypthecodinium cohnii* ; *D. baltica* = *Durinskia baltica*; *K. micrum* = *Karlodinium micrum*; *O. marina* = *Oxyrrhis marina*; *P. shumwayae* = *Pfiesteria shumwayae*.

In the dinoflagellate *Symbiodinium minutum* (order Gymnodiniales) an example of mitochondrial genome expansion could be demonstrated. They contain to date one of the largest mitochondrial genomes (~326 kb) although it contains also only the well-known three protein-coding genes *coxI*, *cox3* and *cob*, as well as 27 fragmented ribosomal RNAs and 12 uncharacterized small RNAs (Shoguchi et al. 2015). The intergenic regions are filled and inflated with non-coding sequences. Together with the 12 small RNAs some of these sequences are similar to those of the malaria agent *P. falciparum* indicating that these non-coding sequences have an unknown function that has been established in a common ancestor of Apicomplexa and Dinoflagellata.

Dinoflagellates which arose by a tertiary endosymbiosis with a stramenopile endosymbiont ('dinotoms') contain mitochondrial genomes that are not reduced in their mitochondrial genome content. These mitochondrial genomes are nearly identical to those of free-living diatoms incorporating more than 30 protein-coding genes (Imanian et al. 2012).

Apicomplexa

The apicomplexan subclass Haemosporida of the Aconoidasida represented by *P. falciparum* has a 6 kb tandemly repeated mitochondrial genome containing the three protein-coding genes *coxI*, *cox3* and *cob*. Additionally, LSU and SSU rRNA genes were found which are highly fragmented into 19 pieces (Preiser et al. 1996; Feagin et al. 1997; Hikosaka et al. 2011). The subclass Piroplasmida (*Theileria* and *Babesia*) has monomeric linear mitochondrial genomes of 6.6 to 8.2 kb with terminal inverted repeats on both ends. These small genomes incorporate also three protein coding genes (*coxI*, *cox3*, *cob*) and six LSU rRNA fragments (Kairo et al. 1994; Lau 2009; Hikosaka et al. 2010). In the Conoidasida (e.g. *Eimeria*) the mitochondrial genome resembles that of *Plasmodium*: a 6.2 kB element arranged in head-to-tail tandem arrays containing the usual three protein-coding genes and 19 rRNA fragments. The structural similarity is independently supported by a phylogenetic analysis (*coxI-cob* concatenated dataset) where *Eimeria* is positioned close to *Plasmodium* indicating that the common ancestor of both genera already contained a concatenated form of the mitochondrial genome (Hikosaka et al. 2011). The mitochondrial genome reconstruction of the Sarcocystidae such as *Toxoplasma gondii* seems to be difficult and to date only multiple copies of partial mitochondrial genes (*coxI* and *cob*) could be identified that are scattered across the nuclear genome (Ossorio et al. 1991; Hikosaka et al. 2011). However, a novel microarray technique allowed to discover the putative mitochondrial genome with a size of ~ 6kb that possibly also

just contains the three standard protein-coding genes of the oxidative phosphorylation pathway (Bahl et al. 2010).

The most extreme genome reduction took place in the early-branching apicomplexan species *Cryptosporidium parvum*. This parasite is lacking a mitochondrial genome and retained a mitosome harboring solely the Fe-S cluster biogenesis pathway (Embley et al. 2003; LaGier et al. 2003).

1.4.5 Complex Plastids in Alveolata

Ciliata

Ciliates are non- photosynthetic strict heterotrophic protists without plastid organelles (Eisen et al. 2006). However, the meanwhile disproven 'Chromalveolate hypothesis' (see above) predicted an plastid ancestry of ciliates followed by a secondary plastid loss and many studies were hence focused on putative plastid-derived genes in contemporary ciliates (Fast et al. 2001; Huang et al. 2004; Hackett et al. 2007; Reyes-Prieto et al. 2008; Archibald 2008). Reyes-Prieto et al. (2008) identified 16 proteins that solidly branch together with Plantae or the CASH lineages (Figure 2). However, ciliates are predators and could acquire such genes from their prey via horizontal gene transfer (HGT) instead of ancient EGTs (Ricard et al. 2006). Some freshwater ciliates harbor algal symbionts of the genus *Chlorella* ('zoochlorellae'), with a strong *Paramecium-Chlorella* species specificity (Summerer et al. 2008). Another example is the kleptoplastidic species *Mesodinium rubrum* (synonym *Myrionecta rubra*), a mixotrophic ciliate causing red tides in the sea (Du Yoo et al. 2015). Several studies showed that *M. rubrum* feeds on cyanobacteria (Du Yoo et al. 2015) and cryptophytes (Yih et al. 2004; Johnson, Stoecker 2005; Park et al. 2007; Hansen et al. 2012) and furthermore retains the undigested bacteria/plastids thus benefitting from their residual photosynthesis. The predatory life style of ciliates facilitates the sporadic recruitment of cyanobacterial and algal genes via HGT, but there is no evidence that this alveolate lineage ever harbored a chloroplast (Petersen et al. 2014).

Dinoflagellata

Most of the photosynthetic dinoflagellates harbor a plastid which is surrounded by three membranes that contains the canonical pigment peridinin alongside with chlorophyll a and c_2 as well as the carotenoid β-carotene (Hackett et al. 2004a). The main curiosity of their plastid genomes is their arrangement in minicircles. Each plastid minicircle (~ 2.2 – 6 kb) incorporates one (Zhang et al. 2002; Mungpakdee et al. 2014) to several (Hiller 2001; Nisbet

et al. 2004; Barbrook et al. 2006) genes and a conserved non-coding core region that is assumed to be the origin of replication (Zhang et al. 1999; Koumandou et al. 2004). To date, the known minicircles harbor two ribosomal genes (16S, 23S rRNA) and 16 plastid protein-coding genes *inter alia* for the photosystem (*psaA*, *psaB*, *psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbI*), the cytochrome b_6f complex (*petB*, *petD*), the ATP synthase complex (*atpA*, *atpB*), the protein translation apparatus (*rpl23*, *rpl28*) and the Fe-S cluster biogenesis (*ycf24* = *sufB*, *ycf16* = *sufC*) (Hackett et al. 2004a). The remaining plastid genes were either lost or transferred to the cell nucleus (Hackett et al. 2004b). Substitutional editing in transcripts of plastid genes is widespread among dinoflagellates (Zauner et al. 2004; Wang, Morse 2006; Dang, Green 2009; Jackson, Waller 2013; Mungpakdee et al. 2014).

Dinoflagellates are characterized as the 'kings of symbioses' (Morden, Sherwood 2002), thus they often underwent plastid replacements and additional serial endosymbioses. The two membrane-bound plastid of *Lepidodinium viride* as well as *Gymnodinium chlorophorum* (order: Gymnodiniales) contains the pigment prasinoxanthin as its main photopigment indicating a dinoflagellate plastid replacement by a green algal prasinophyte symbiont (Saldarriaga et al. 2001). Tertiary endosymbioses were demonstrated in the 'dinotoms' *Kryptoperidinium foliaceum* and *Durinskia baltica* which arose by the assimiliation of a diatom endosymbiont (Chesnick et al. 1997), whereas the marine species *Karlodinium veneficum*, *Gymnodinium aureolum* and *Karenia brevis* harbor the pigment 19'-hexanoyloxy-fucoxanthin instead of perdinin in their plastids – an indicator for a tertiary endosymbiosis with a haptophyte endosymbiont (Tengs et al. 2000).

Kleptoplastidy is likewise widely distributed among dinoflagellates. Thus the fresh water species *Gymnodinium acidotum* and the marine species *Amphidinium wigrense* acquire their phototrophy by transient plastids originating from cryptophytes (Wilcox, Wedemayer 1985; Xia et al. 2013). The marine mixotrophic species *Dinophysis acuminata* even steels itself the cryptophycean kleptoplastid of the ciliate *M. rubrum* (Takishita et al. 2002; Wisecaver, Hackett 2010; previous subchapter).

In the earliest-branching dinoflagellate genus *Perkinsus* there is a clear evidence of nuclear-encoded plastid proteins, but the final proof for the presence of a plastid is missing (Grauvogel et al. 2007; Stelter et al. 2007; Matsuzaki et al. 2008; Joseph et al. 2010; Fernandez Robledo et al. 2011). However, a complex plastid surrounded by four membranes was proposed based on electron microscopic pictures of the species *Perkinsus olseni* (Teles-Grilo et al. 2007). The same seems to be true for the non-photosynthetic heterotrophic early-branching dinoflagellate *O. marina*, which seems to harbor a cryptic plastid or harbored once

a photosynthetic active plastid by showing plastid-derived genes in a large-scale EST dataset (Slamovits, Keeling 2008).

Apicomplexa

In 1996 Wilson et al. discovered a heterotrophic plastid in the malaria parasite *P. falciparum*, thus documenting that the ancestor of all Apicomplexa once acquired a photosynthetic plastid by endosymbiosis (Ginger 2006; Wilson et al. 1996). There is ultrastructural evidence that this organelle has been acquired by a secondary endosymbiosis event, because it is surrounded by four membranes: the former double-membrane of the endosymbiont's plastid, the plasma membrane of the endosymbiont itself and the membrane derived from the food vacuole, where the endosymbiont was once absorbed by the heterotrophic host cell (Ginger 2006; Waller et al. 2000). Since the plastid in Apicomplexa has lost the ability of photosynthesis due to their parasitic lifestyle, one cannot refer to it as a 'chloroplast' in an original sense. That is why plastids of Apicomplexa are designated as apicoplasts (McFadden 2011).

The phylum Apicomplexa also includes some parasitic lineages which do not incorporate a plastid: (1) the species of the parasitic genus *Cryptosporidium* are supposed to have lost the plastid genome and the entire organelle (Zhu et al. 2000; Abrahamsen et al. 2004; Xu et al. 2004), because some relict nuclear-encoded and expressed plastid genes are still present (Huang et al. 2004); (2) the non-photosynthetic apicomplexan group of gregarines might also lost both, the plastome and the plastid itself (Toso, Omoto 2007). However, after the refutation of the 'Chromalveolate Hypothesis' (Baurain et al. 2010; Petersen et al. 2014) the common plastid ancestry of all Apicomplexa has not been documented yet and a definite proof of plastid loss in both lineages is thus still missing.

In recent years, the discovery of the two photosynthetic algae *C. velia* and *V. brassicaformis* which are closely related to the parasitic apicomplexans, revolutionized the research perspectives for plastid evolution in alveolates (Moore et al. 2008; Obornik et al. 2012). Their chloroplast is surrounded by four membranes and the pigmentation is indicative of an origin by secondary endosymbiosis with a red alga (Keeling 2004). Furthermore, they share the common feature of a non-canonical UGA-codon for the amino acid tryptophan with the apicoplast (Moore et al. 2008; Obornik et al. 2012). It is likely that the contemporary chromerid plastid is very similar to the ancestral plastid of all Apicomplexa and accordingly represents an ideal reference system to study its cell biology and evolution.

1.4.6 Peroxisomes in Alveolata

Peroxisomes were first described and extensively studied in ciliates and *Tetrahymena pyriformis* was even a reference organism for the introduction of the term 'peroxisome' (Hogg, Kornberg 1963; Duve, Baudhuin 1966; Seaman 1970). Their peroxisomes are considered as 'glyoxysomes' incorporating the eponymous glyoxylate cycle in which ciliates are able to use acetyl-CoA as a carbon source which is obtained from the degradation of fatty acids in order to synthesize carbohydrates (Kornberg, Krebs 1957; McCammon et al. 1990; Schnarrenberger, Martin 2002).

For a long time little was known about peroxisomes in the other alveolate groups leading to the assumption that they are absent in Apicomplexa and Dinoflagellata (reviewed in Gabaldon 2010). In dinoflagellates, the presence of 'microbodies' was predicted based on ultrastructural pictures (Bibby, Dodge 1973) and only recently the discovery of transcripts encoding for two genes of the glyoxylate cycle provides an independent hint for the presence of peroxisomes in dinoflagellates (Butterfield et al. 2013).

A comparative *in silico* analysis of eukaryotic genomes revealed that Apicomplexa are most likely lacking peroxisomes in the presence of full functional mitochondria (Schlüter et al. 2006). However, the presence of a peroxisome organelle was hypothesized for the feline apicomplexan pathogen *T. gondii* by the immunofluorescence staining of the peroxisomal catalase protein in a vesicular compartment (Kaasch, Joiner 2000), but markers for the biogenesis of the peroxisome and for other metabolic pathways seem to be absent (Ding et al. 2000). Searches for the two glyoxylate genes found in dinoflagellates within completely sequenced apicomplexan lineages such as *Toxoplasma* and *Plasmodium* revealed no results (Butterfield et al. 2013). Thus, the existence and distribution of peroxisomes in this alveolate phylum remains cryptic and is still under debate (Gabaldon et al. 2016).

1.5 Essential Organelle Functions

1.5.1 Mitochondrial Pathways

Mitochondria are involved in a set of metabolic pathways (Danne et al. 2013) but it is conspicuous that in all reduced mitochondrial genomes of alveolates, the genes for the electron transport chain (ETC) of the oxidative phosphorylation (OXPHOS) step are retained (coxI, cox3, cob; chapter 1.4.4). The iron-sulfur cluster synthesis is an essential mitochondrial

function and seems to be the only function retained in MROs (Lill, Kispal 2000; Embley et al. 2003; Shiflett, Johnson 2010; Diekmann, Pereira-Leal 2013).

Oxidative Phosphorylation

This pathway is also known as the 'respiratory chain' of the cell where energy in form of adenosine triphosphate (ATP) is generated by a membrane gradient of protons. On the other hand, this gradient is produced by an electron transport chain (ETC) across diverse protein complexes within the mitochondrial cristae (Chaban et al. 2014; Figure 4). Dysfunctions and the release of reactive oxygen species (ROS) in these protein complexes are associated with aging (Balaban et al. 2005) and neurodegenerative disorders in humans such as Parkinson or Alzheimer's disease (Kim et al. 2000; Bender et al. 2006; Hroudova et al. 2014).

The first protein complex (**complex I**, Figure 4) is the NADH dehydrogenase or NADH-coenzyme Q oxidoreductase and one of the largest protein assemblies (~980 kDa in eukaryotes) in this energy-collecting pathway. Complex I has an L-shape with a hydrophobic domain embedded in the mitochondrial membrane and a hydrophilic domain positioned in the mitochondrial matrix. Flavin mononucleotides (FMN) and cascades of Fe-S clusters serve as prosthetic groups within the ETC which uses the NADH to reduce the coenzyme Q10 (Q) to ubiquinol. At the same time, four protons per NADH are pumped into the intermembrane space (Efremov et al. 2010; Figure 4).

In contrast, **complex II** is not involved in the proton gradient formation. This succinate dehydrogenase originated from the tricarboxylic acid cycle (TCA) and was integrated into this pathway. It catalyzes the reversible oxidoreduction of succinate to fumarate and transfers the incoming electrons with the help of the coenzyme flavin adenine dinucleotide (FAD) and Fe-S clusters to Q which is subsequently reduced to ubiquinol (Iverson 2013; Figure 4).

During the oxidation of ubiquinol (QH_2), one cytochrome c molecule per electron is reduced and two protons are transmitted into the intermembrane space. Hence, the protein assembly **complex III** (Figure 4) is a cytochrome c oxidoreductase. A catalytic core is located within this protein assembly to ensure the electron transport to the cytochrome c molecule. This catalytic core is composed of the three subunits *cyt* c_1, *cyt-b* and the Fe-S cluster of the Rieske iron-sulphur protein (Crofts 2004). The reaction within complex III proceeds in two steps based on the so-called Q-cycle. (1) Because only one electron at a time can be accepted by cytochrome c, the second electron from the oxidation of QH_2 is absorbed by the substrate Q which is transformed into the free radical Q^- which resides for the moment. (2) Subsequently a second ubiquinol molecule is oxidized leaving again two electrons out of this reaction. The first electron is transported via the catalytic core to the bound cytochrome c, whereas the

second electron is submitted to the free radical Q^- which is immediately reduced to QH_2 consuming two protons from the mitochondrial matrix site (Trumpower 1990).
The final protein complex of the ETC is the Cytochrome c oxidase (**complex IV**, Figure 4). This transmembrane protein complex contains heme groups and metal ions as cofactors to transmit the electrons resulting from the oxidation of the bound cytochrome c to oxygen which is in turn reduced to water (Tsukihara et al. 1996). Chemical protons from the mitochondrial matrix are assigned to this water formation whereas the energy out of this reaction path is used to pump two protons into the intermembrane space (Calhoun et al. 1994).
The constructed proton gradient of the three protein complexes (complex I, III and IV) is utilized by the enzyme complex ATP-synthase (**complex V**, Figure 4) to generate energy for the cell. This protein complex has an asymmetric structure composed of a stalk and a knob. Furthermore the protein complex is compartmentalized into the membrane-bound F_0 portion involved in proton translocation and the soluble F_1 portion responsible for the ATP-hydrolysis. Both, stalk- and knob-region are made up of these two portions. Three pumped protons are translocated through a channel-like structure from the intermembrane space to the mitochondrial matrix to be processed for one ATP molecule (Boyer 1997; Figure 4).

Among alveolates ciliates harbor the conventional oxidative phosphorylation pathway consisting of complex I, II, III and IV (Danne et al. 2013). Surprisingly, the *cox3* gene of complex IV is missing from the mitochondrial genome of ciliates (Burger et al. 2000; Brunk 2003; Graaf et al. 2009) and could even not be found as a nuclear-encoded mitochondrial gene in a macronuclear genome sequence of *T. thermophila* (Eisen et al. 2006). Additionally, the ciliates show a highly divergent ATP-synthase complex incorporating 13 novel proteins which were identified by a proteome analysis (Balabaskaran et al. 2010; Chaban et al. 2014).
In a common ancestor of Apicomplexa and dinoflagellates (‘Myzozoa’, Cavalier-Smith, Chao 2004), the whole protein complex I of the mitochondrial ETC got lost (Gardner et al. 2002; Adams, Palmer 2003; van Dooren et al. 2006; Waller, Jackson 2009; Danne et al. 2013). The loss of the protein complex I has also been shown before for some yeasts such as *S. cerevisiae* (Vries, Grivell 1988; van Dooren et al. 2006). However, an alternative NADH dehydrogenase similar to those found in fungi and some land plants (Kerscher 2000) was discovered in apicomplexan as well as in dinoflagellate genomes indicating that this enzyme could replace the initial complex I of the oxidative phosphorylation for the reduction of coenzyme Q (Danne et al. 2013). This enzyme has already been identified in *Toxoplasma* and it is located

in the mitochondrial matrix (Lin et al. 2011). Ciliates do not contain such an enzyme, thus it must have been established in the common ancestor of the Myzozoa (Danne et al. 2013).

A synapomorphy shared by the unrelated groups Myzozoa and Chlorophyceae is the splitting of the *cox2* gene of complex IV into two parts (*cox2a* and *cox2b*) which are encoded in the nucleus (Waller, Keeling 2006). This special feature provoked the discussion of a green algal ancestry of the apicoplast (Funes et al. 2002; Waller et al. 2003). However, this discussion was finally refused as phylogenetic reconstructions of the apicoplast clearly support a red algal origin (Janouskovec et al. 2010) and as horizontal gene transfer in protist genomes occur more often than expected like shown for the 'green phosphoribulokinase (PRK)' in CASH lineages. This gene was recruited from a green alga and replaced the red algal equivalent (Petersen et al. 2006). Interestingly, in ciliates the COX2 protein is encoded on the mitochondrial genome and has a 300 amino acid (aa) insertion at the position where the myzozoan *cox2* was split indicating the plasticity of this gene (Waller et al. 2003).

Apicomplexan parasites often harbor an alternative oxidase (AOX) instead of (*C. parvum*) or additionally to the whole oxidative phosphorylation pathway (*P. falciparum*, *T. gondii*; Roberts et al. 2004). This enzyme bypasses complex III guiding the electrons directly to oxygen after the ubiquinone complex (McIntosh 1994).

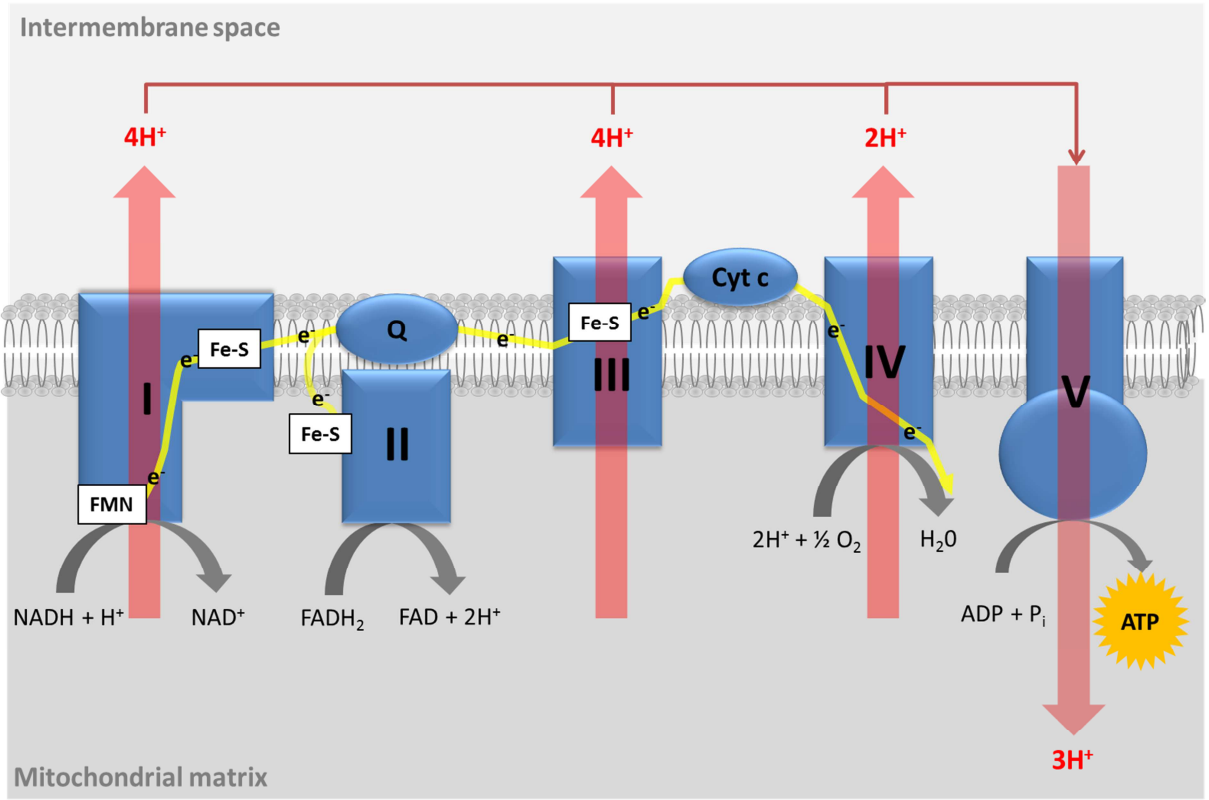


Figure 4: Schematic representation of the oxidative phosphorylation process for energy production in mitochondria. The ATP-generation is interconnected with an electrochemical gradient which is synthesized by a cascade of chemical reactions catalyzed by three different protein complexes (complex I, III, IV). A running electron transport chain through all proteins and protein complexes initiates the pumping of protons from the mitochondrial matrix to the intermembrane space. This process is also called 'cell respiration' because oxygen is used as the terminal electron acceptor to produce water. Abbreviations are in compliance with the text. H^+ = Proton; e^- = Electron.

Iron-sulphur Cluster Assembly (Isc System)

Fe-S clusters are important cofactors in chemical reactions and electron transport accordingly they are likewise present in the catalytic cores of the protein complexes of the oxidative phosphorylation (Lill, Mühlenhoff 2006; Danne et al. 2013; Figure 4). The proteins for Fe-S protein biogenesis in mitochondria are closely related to the bacterial isc operon for Fe-S cluster assembly, hence indicating that this system was adopted from the α-proteobacterial endosymbiont (Tovar et al. 2003; Lill 2009; Danne et al. 2013). The Isc system can be divided into two steps: (1) An Fe-S cluster is temporarily assembled at the scaffold proteins **IscU** and **IscA** depending on (2) iron and sulphur donors. The cysteine desulphurase **IscS** converts cysteine to alanine in order to release sulphur. The sulphur is bound in form of persulphide in cysteine residues of the IscS-enyzme or helper proteins to be transported to the scaffold protein (Lill 2009). The iron donor to form the Fe-S clusters is still under debate, but there are indications that on the one hand the Fe^{2+}-ions come from the protein Frataxin which is a nuclear-encoded mitochondria-located protein for iron homeostasis (Yoon, Cowan 2003) or these ions are received by Glutathione which is involved in the antioxidant system of cells and serves as a redox buffering agent (Qi et al. 2012).

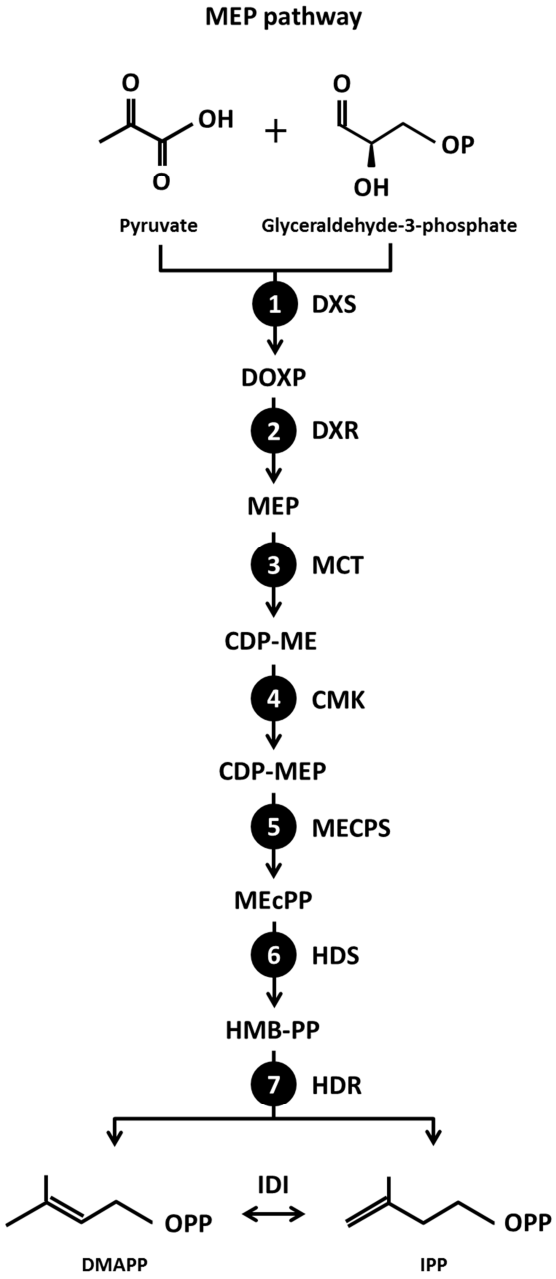
An electron transfer is needed to convert the cysteine-residue-bound S^0 to sulphide (S^{2-}) which is needed for the Fe-S clusters. This electron transfer can be accomplished by Ferrodoxin (Fdx) or a ferrodoxin-$NADP^+$ reductase (mtFNR). Subsequently both substrates, sulphide as well as Fe^{2+}, are assembled to Fe-S clusters at the scaffold protein IscU/IscA. These generated Fe-S clusters can be transferred to apoproteins which are immediately converted into their holo forms (Lill 2009).

In all three alveolate phyla, the proteins for the generation of Fe-S clusters via the Isc system in mitochondria are present (Danne et al. 2013). Even in the amitochondriate apicomplexan genus Cryptosporidium where an organelle-like mitosome retained only this pathway (LaGier et al. 2003).

1.5.2 Plastid Pathways

Despite the loss of photosynthesis, the alveolate parasites depend on their plastids by de novo synthesis of crucial chemical compounds such as isoprenoids (the 2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate pathway; MEP/DOXP pathway), iron-sulfur clusters (Suf system), tetrapyrroles (C5 pathway) and fatty acids (FASII) (Ralph et al. 2004; Vaughan et al. 2009; Seeber, Soldati-Favre 2010; Yeh, DeRisi 2011; Ramakrishnan et al. 2012; Gisselberg et al. 2013; Gornik et al. 2015).

MEP Pathway (or 'Non-mevalonate Pathway')



This metabolic pathway is responsible for the production of two precursors of isoprenoids, dimethylallyl pyrophosphate (DMAPP) and isopentenyl pyrophosphate (IPP), which are isomers constantly interconverted by the enzyme isopentenyl pyrophosphate isomerase (IDI). These precursors are used to construct more complex structures such as isoprenoids, steroids, lipids or terpenes (e.g. important for carotinoid production) in further reactions (Rohmer et al. 1993; Lange et al. 2000; Cordoba et al. 2009; Hunter 2007).

The MEP pathway (Hunter 2007; Figure 5) consists of a cascade of seven series connected enzymatic reactions and its first step depending on glyceraldehyde-3-phosphate and pyruvate is linked with the glycolysis in the cytoplasma (Furumoto et al. 2011). The metabolite glyceraldehyde-3-phosphate as well as the product pyruvate are the initiators of the whole pathway. The condensation of these two reactants to 1-desoxy-D-xylulose 5-phosphate (DOXP) is realized by the first enzyme 1-desoxy-D-xylulose-phosphate synthase (**DXS**, EC 2.2.1.7) and the product is subsequently interconverted by the enzyme 1-deoxy-D-xylulose 5-phosphate

Figure 5: Schematic representation of the enzymatic steps in the MEP pathway for isoprenoid precursor synthesis. Abbreviations are according to the text.

reductoisomerase (**DXR**, EC 1.1.1.267) into 2-C-methyl-D-erythritol 4-phosphate (MEP). In the third step the transferase **MCT** (2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase, EC 2.7.7.60) converts MEP to 4-diphosphocytidyl-2C-methyl-D-erithritol (CDP-ME) with the help of Cytidine triphosphate (CTP). The disposal of ATP within the reaction of the fourth MEP enzyme 4-diphosphocytidyl-2-C-methyl-D-erithritol kinase (**CMK**, EC 2.7.1.148) leads to the synthesis of 4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate (CDP-MEP). Afterwards a cyclization takes place: cytidine monophosphate is splitted off to generate a phosphate ester bond between C_2 and C_4 for the next intermediate product 2-C-methyl-D-erythritol 2,4-cyclodiphosphate (MEcPP). This step is catalyzed by the enzyme 2-C-methyl-D-erythritol 2,4-cyclodisphosphate synthase (**MECPS**, EC 4.6.1.12). The 4-hydroxy-3-methylbut-2-en-1-yl disphosphate synthase (**HDS**, EC 1.17.7.1) converts MEcPP to 4-hydroxy-3-methyl-but-2-enyl pyrophosphate (HMB-PP) using reduced ferrodoxins (4Fe-4S clusters) as reduction equivalents and splitting off water. The final reaction is catalyzed by the iron-sulfur protein 4-hydroxy-3-methylbut-2-enyl diphosphate reductase (**HDR**, EC 1.17.7.4) incorporating Fe-S clusters as cofactors. By means of the reductive equivalent NAD(P)H HMB-PP is transformed into the products IPP and DMAPP whose interconvention by the isopentenyl diphosphate isomerase (**IDI**, EC 5.3.3.2) results in a metabolite ratio of approximately 5:1 (Rohdich et al. 2002).

The MEP pathway is present in all plastid-bearing alveolates (Gile, Slamovits 2014). The absence of this pathway in humans made it a useful target in the development of drugs against important human pathogens (Lichtenthaler 2000; Zeidler et al. 2000; Rohdich et al. 2002; Hunter 2011). Thus, in *Plasmodium falciparum* one protein of this pathway had surprisingly been identified as a useful drug target against this parasite indicating that the functional pathway for isoprenoid synthesis in the apicoplast is essential. For instance, by inhibiting the protein DXR with the herbicide fosmidomycin, the human pathogen could not survive (Jomaa et al. 1999). Recently, it could be shown that these fosmidomycin-treated blood stage *P. falciparum* cells can be rescued from growth inhibition by the supplementation of the downstream isoprenoid precursor IPP (Yeh, DeRisi 2011). As a consequence, *P. falciparum* cells cultivated with IPP and additional antibiotics could be shown to lose their apicoplast genome over multiple cell cycles as the protein import system got lost. Accordingly, these cells harbor a non-intact apicoplast, thus they are completely dependent on the exogenous IPP (Yeh, DeRisi 2011).

Iron-sulphur Cluster Biosynthesis (Suf Pathway)

The Suf pathway is the most ancient biogenesis pathway for iron-sulfur (Fe-S) clusters and was likely already present in early anaerobic life forms. Because of their high reactivity, Fe-S clusters are important cofactors for electron transfer as well as substrate binding and activation reactions (Boyd et al. 2014). In some organisms such as *P. falciparum* the Suf pathway is the only system present for the biogenesis of Fe-S clusters and it is therefore essential for the viability of the respective organism (Gisselberg et al. 2013). However, it often works in parallel with the mitochondrial Isc- and the cytosolic Nif-system (Lill 2009; Boyd et al. 2014). The MEP pathway for isoprenoid synthesis is dependent on the biosynthesis of Fe-S clusters that is represented by the Fe-S cluster – relying proteins HDS and HDR (see subchapter 'MEP Pathway'; Dellibovi-Ragheb et al. 2013). The Suf machinery is complementary to the Fe-S cluster biogenesis in cyanobacteria indicating that plastids retained this ancient pathway from their former endosymbionts (Lill 2009; Boyd et al. 2014).

In comparison to the Isc system (chapter 1.5.1), the Suf system also consists of two steps representing iron- and sulphur-acquisition as well as the assembly of these ions to Fe-S clusters. The initiation step is performed by a cysteine desulphurase (**SufS**, EC 2.8.1.7) which transfers the sulphur to a cysteine residue of the assisting protein **SufE**. As in the Isc system the iron donor as well as the electron donor is still unknown. However, the *de novo* formation of Fe-S cluster is mediated by three different scaffold proteins: **SufU**, **SufA** and the protein complex **SufBCD** whereas SufC represents an ATPase which might facilitate the dissociation of Fe-S clusters from SufB and the transfer to apoproteins (Lill 2009). The presence of the full SufABCDSE system was already shown for the malaria parasite *P. falciparum* (Outten 2015). However, the Suf pathway is presumed to be also present in other apicomplexan genera such as *Toxoplasma*, *Babesia* and *Theileria* (Seeber, Soldati-Favre 2010). Additionally, *sufB* could be identified in the plastid genome of *V. brassicaformis* but not in *C. velia* (Janouskovec et al. 2010). Recently, it was discovered that blood stage and liver stage cells of *Plasmodium* can be inhibited by the antibiotic D-cycloserine which inhibits the desulphurase activity of SufS and SufE (Charan et al. 2014). This example documents the importance of this plastid-specific pathway for the viability of the parasite.

In dinoflagellates the Suf system is not well researched, because the genes and proteins of this pathway were often assigned to hypothetical proteins of the classification *ycf* (hypothetical chloroplast open reading frame). Anyway, plastidial minicircles containing *ycf24* (SufB) or *ycf16* (SufC) were described for the dinoflagellate species *Ceratium horridum* (Laatsch et al. 2004), although they could represent an independent acquisition by bacterial horizontal gene

transfer (Moszczynski et al. 2012). However, the peridinin-containing dinoflagellate *Amphidinium carterae* shows also a *sufC*-gene which was transferred from the chloroplast to the host nucleus (Bachvaroff et al. 2014).

Tetrapyrrol Biosynthesis

The biosynthesis of tetrapyrrol is the prerequisite to generate chlorophyll in photoautotrophic organisms or to produce heme for the oxidative and energy metabolism (Koreny et al. 2011). There exist two different ways of how to produce tetrapyrroles: the C5 pathway and the C4 pathway. Most photosynthetic organisms use the C5 pathway which is strictly located in the lumen of the plastid. The whole C5 pathway includes a cascade of nine proteins (Tanaka, Tanaka 2007; Figure 6). The first precursor in tetrapyrrole biosynthesis is 5-aminolevulinic acid (ALA) which is synthesized from glutamyl-tRNA by glutamyl-tRNA reductase (**Gr**; Gornik et al. 2015) and the enzyme glutamate 1-semialdehyde aminotransferase (**Ga**; Gornik et al. 2015).

Subsequently, ALA is converted to tetrapyrroles by seven enzymatic reactions consisting of the porphobilinogen synthase (**HemB**), porphobilinogen deaminase (**HemC**), uroporphyrinogen III synthase (**HemD**) and urophorphyrinogen III decarboxylase (**HemE**), coproporphyrinogen oxidase (**HemF**), protoporphyrinogen oxidase (**HemG/Y**) and ferrochelatase (**HemH**). The C5 pathway in photosynthetic eukaryotes was once obtained by the cyanobacterial endosymbiont and thus retained in the plastid organelle. In contrast the 'ancestral' C4 pathway (Figure 6), which is abundant in heterotrophic eukaryotes, is located in both the mitochondrion and the cytosol. The first precursor ALA is synthesized from the condensation between glycine and succinyl-CoA catalyzed by the ALA synthase (**ALAS**) in the mitochondrion. ALA is then exported to the cytosol where the four enzymes HemB, HemC, HemD and HemE convert it to coproporphyrinogen. This reaction product is again transported into the mitochondrion where it is transformed by the enzymes HemF, HemG/Y and HemH to the tetrapyrrole protoheme (van Dooren et al. 2012; Gornik et al. 2015).

The majority of photosynthetic dinoflagellates use the C5 pathway for tetrapyrrole synthesis. However, in early branching heterotrophic lineages such as *Perkinsus*, *Oxyrrhis* and *Hematodinium*, the C4 pathway enzyme ALAS was detected corresponding to their non-photosynthetic lifestyle. Further investigations in *Hematodinium* sp. revealed a mosaic pathway of both C4 and C5 pathway with plastid-derived C5 enzymes located in the cytosol of the cell. The composition is comparable to that of apicomplexan parasites that also harbor a hybrid of both tetrapyrrole pathways. This discovery implies that both pathways must have

been present in a common ancestor of the Myzozoa (Figure 6). The hypothesis is supported by the presence of the hybrid pathway version in the photosynthetic apicomplexan alga *C. velia*, which thus represents a connecting link between parasitic Apicomplexa and dinoflagellates. This unusal pathway starts with the production of ALA by ALAS in the mitochondrion, which is subsequently transported into the plastid. In *Chromera* all next enzymatic steps take place in the plastid, whereas in apicomplexan parasites only a part of these conversions is performed in the plastid (3 to 4 enzymatic steps) and the final three reactions occur again in the mitochondria (Figure 6; Koreny et al. 2011; van Dooren et al. 2012; Gornik et al. 2015).

Fatty Acid Biosynthesis (FASII)

Fatty acids are the most important components of cellular membranes and they are used as a source to store energy (Seeber, Soldati-Favre 2010). In addition or in exchange to the cytosolic multienzyme type I FAS complex (homologous to polyketide synthases), fatty acids can also be synthesized in the plastid lumen by a cascade of seven separate monofunctional proteins of the type II FAS pathway, which was once acquired by the cyanobacterial symbiont (Gornik et al. 2015; Figure 6).

A comprehensive screening approach in dinoflagellates revealed that the type I FAS as well as type II FAS genes are omnipresent in photosynthetic species (Kohli et al. 2016). In heterotrophic species such as *Hematodinium* sp. or *Crypthecodinium cohnii* it was shown that only the cytosolic FAS I pathway is present (Sonnenborn, Kunau 1982; van Dolah et al. 2013; Gornik et al. 2015). Surprisingly, FAS II genes have also been identified in the heterotrophic dinoflagellate *Oxyrrhis marina* (Kohli et al. 2016), thus providing evidence for the presence of a formerly overlooked plastid (Slamovits, Keeling 2008).

In apicomplexan parasites the distribution of the FAS I and FAS II pathways is very heterogeneous. Thus, while the genera *Toxoplasma, Neospora* and *Eimeria* harbor the cytosolic as well as the plastidic pathway, *Plasmodium* incorporates only the type II form in the apicoplast. The aplastidic genus *Cryptosporidium* harbors only the cytosolic pathway as expected (Zhu 2004; Seeber, Soldati-Favre 2010; Kohli et al. 2016).

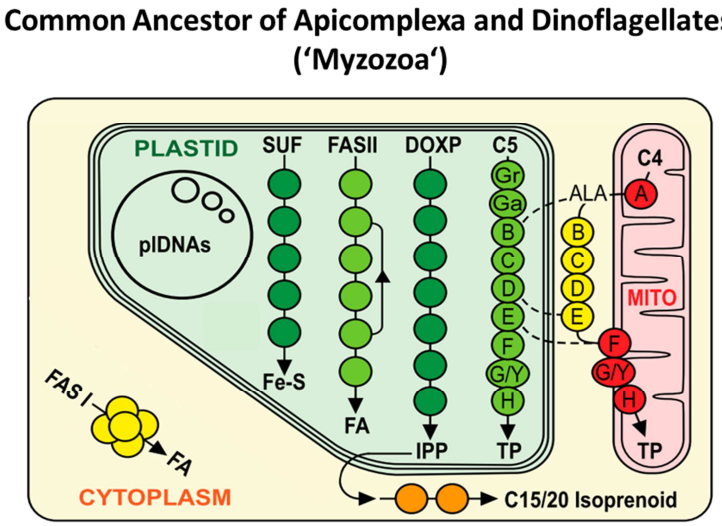


Figure 6: Reconstructed plastid relevant metabolic pathways in the common ancestor of apicomplexans and dinoflagellates (according to Gornik et al. 2015). Abbreviations: DOXP = 1-deoxy-D-xylulose 5-phosphate pathway (MEP pathway); FAS = Fatty acid synthesis; Fe-S = iron-sulfur cluster; IPP = isopentenyl pyrophosphate; plDNAs = dinoflagellate plastid minicircles; SUF = Iron-sulfur cluster assembly;TP = tetrapyrroles. Protein abbreviations are according to the text.

1.5.3 Peroxisomal Biogenesis and Metabolic Pathways

The peroxisome-specific biogenesis proteins (peroxins) can be assigned to specific functional categories. To date, 34 different peroxins are known whereupon some peroxins are only found in fungi, plants or mammals (see: PeroxisomeDB; Schlüter et al. 2010). 16 predominantly occurring peroxins are discussed in the following chapter (see: KEGG pathway database; Kanehisa, Goto 2000). Furthermore, I compile the most important metabolic pathways of peroxisomes in protists.

Peroxisomal Biogenesis

Peroxisomal matrix proteins carry particular peroxisomal targeting sequences (PTS) which are recognized by the '**importomer**' (Figure 7) of the protein import system I. The cargo receptors Pex5 and Pex7 detect the C-terminal PTS (PTS1) or N-terminal PTS (PTS2), respectively. The matrix protein is temporarily bound to the receptors, which are forming a complex with a transmembrane docking station Pex13 and/or Pex14 (Williams, Distel 2006). It is assumed that the matrix proteins are subsequently conveyed through a transient pore into the matrix of the organelle (Meinecke et al. 2010).

After the release of its cargo, Pex5 can be dislocated via the ATP-dependent '**exportomer**' (Figure 7) into the cytosol. On the one hand, Pex5 can enter a recycling path by being monoubiquitinated via the ubiquitin-conjugating enzyme Pex4 anchored by Pex22. The ubiquitin is transferred with the assistance of the RING-Finger domain proteins and ubiquitin

ligases Pex2, Pex10 and Pex12. The 'marked' Pex5 is exported to the cytosol by the AAA-type ATPase complex Pex1/Pex6 linked to its membrane anchor Pex15 (Pex26 in humans). In the cytosol the bound ubiquitin of Pex5 is released and the receptor can enter a new cycle of matrix protein import. Another strategy incorporates the polyubiquitination of the receptor Pex5 again with the help of the three RING finger proteins. This construct is also dependent on the Pex1/Pex6 complex for its transport into the cytosol where its 'poly-label' is a signal for the 26S proteasome to degrade Pex5 (Platta et al. 2013). The energy-dependent export of Pex5 by the Pex1/Pex6 complex is hypothesized to trigger the importomer ('export-driven import hypothesis'; Schliebs et al. 2010). Both proteins of this ATPase complex are closely related to the ATPase of the ERAD system, indicating that both machineries share a common evolutionary origin (Bolte et al. 2011).

A further cargo receptor, Pex19, is responsible for the translocation of integral peroxisomal membrane proteins (PMPs) within the protein import system II that are responsible for the **membrane assembly** (Figure 7) of the peroxisome. Pex19 interacts with hydrophobic domains within the PMPs, the membrane PTSs. Thereby the receptor stabilizes the PMPs like a chaperone and directs them to the peroxisomal membrane (Jones et al. 2004). The transmembrane docking partner for Pex19 and its cargo is Pex3 (Fang et al. 2004), whereas in mammalian peroxisomes it is bound by the transmembrane protein Pex16 (Honsho et al. 2002). A peroxisomal integral membrane protein, Pex11, which is also imported by the Pex19-dependent protein import system, mediates the elongation of peroxisome before dynamin-related proteins intitiate the fission process in **peroxisomal organization** (Fagarasanu et al. 2007; Figure 7).

Peroxisomal Pathways

Numerous organisms harbor the anabolic **glyoxylate cycle** (Figure S1, red), a bypass of the mitochondrial TCA cycle in order to produce the C4 unit succinate from the carbon source and the C2 unit acetyl-CoA. The cycle is composed of five successive enzymatic reactions performed by the aconitase (ACO), isocitrate lyase (ICL), malate synthase (MLS), malate dehydrogenase (MDH) and citrate synthase (CS). The different steps are localized on both sides of the peroxisomal membrane, the cytosol and the peroxisomal matrix. In land plants and fungi, the oxidation of malate to oxaloacetate by the MDH and the interconversion of citrate to isocitrate by ACO, resides in the cytosol. The remaining enzymes ICL, MLS and CS are imported into the peroxisomal matrix.

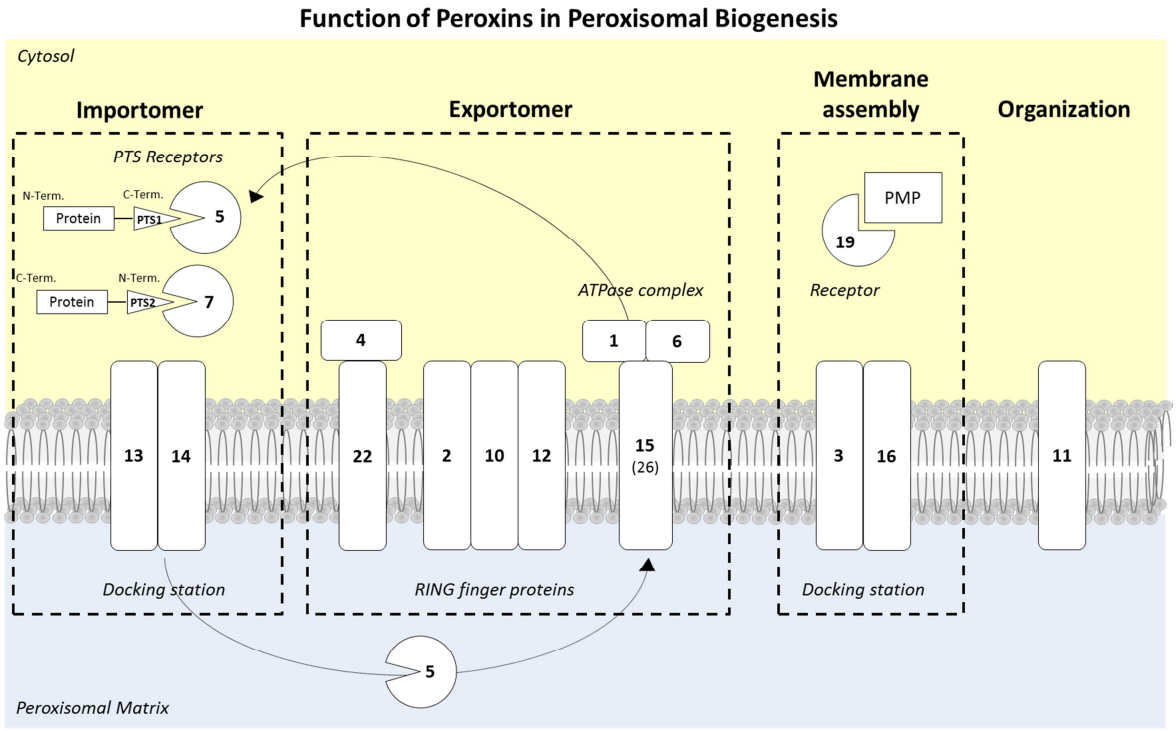


Figure 7: Schematic overview of the function of peroxins in peroxisomal biogenesis. Shown are the four functional categories: 'importomer', 'exportomer', 'membrane assembly' and 'organization' taking place in the cytosol (yellow) as well as in the peroxisomal membrane and matrix (grey). Abbreviations are according to the text.

The reaction intermediates are exchanged in a ping-pong-mechanism via to date unknown transporters across the peroxisomal membrane (question marks, Figure S1; Kunze, Hartig 2013). Three isoenzymes of the glyoxylate cycle (ACO, CS and MDH) are also essential for the mitochondrial TCA cycle (Schnarrenberger, Martin 2002) and additional isoforms of MDH are required for the plastid energy metabolism (Scheibe 2004). However, the ICL and MLS are classified as peroxisome-specific key markers for the glyoxylate cycle and were often used for the identification of peroxisomes (Tolbert, Essner 1981; Lazarow, Fujiki 1985; Kunze et al. 2006; Butterfield et al. 2013).

Another anabolic pathway is the **biosynthesis of ether phospholipids** (Figure S1, blue) which are a key component of the cell membranes. The pathway is composed of three enzymatic steps. The initial step is accomplished by dihydroxyacetone phosphate acyltransferase (DHAPAT) which acylates the substrate dihydroxyacetone phosphate (DHAP) with an acyl-CoA. The next step requires fatty alcohols which are synthesized by the enzyme fatty acyl reductase (FAR) from acyl-CoAs/fatty acids. These primary fatty acid alcohols are used by the enzyme alkyldihydroxyacetone phosphate synthase (AGPS) to replace the acyl-group of DHAP by an alkyl-group in order to generate ether bonds. Interestingly, in the ciliate species *T. thermophila* it was found a bifunctional protein 'TtFARAT' originated by a gene

fusion of the N-terminal sequence for the FAR and the C-terminal sequence for the DHAPAT (Dittrich-Domergue et al. 2014).

The **β-oxidation of fatty acids** (Figure S1, yellow) is a catabolic process by degrading fatty acid molecules via the oxidation of their beta carbon to a carbonyl group generating acetyl-CoA. Whereas in mitochondria only short-chain fatty acids enter the fatty acid β-oxidation pathway, in peroxisomes very-long-chain fatty acids are prepared before they are sent to the mitochondrial pathway. Such fatty acids include molecules with a chain length of 24 to 26 carbon-atoms per fatty acid, methyl-branched chain fatty acids and intermediates of the bile acid synthesis pathway. Fatty acids containing a methyl group such as phytanic acid are processed by the α-oxidation pathway incorporating the enzymes phytanoyl-CoA hydroxylase (PHYH) and 2-hydroxyacyl-CoA lyase 1 (HPCL2) before entering the β-oxidation step. The additional enzyme alpha-methylacyl-CoA racemase (AMACR) converts (2R)-methyl fatty acyl-CoAs into (2S)-methyl fatty acyl-CoAs which is an obligatory method for the oxidation of (2R)-fatty acids. The β-oxidation of fatty acids is organized in dehydrogenation, hydration, dehydrogenation and finally the thiolytic cleavage. The initial step is accomplished by the flavoprotein acetyl-CoA oxidase (ACOX) which produces hydrogen peroxide as a byproduct. The next steps are realized by a bifunctional protein including an enoyl-CoA hydratase and 3-hydroxy-acyl-CoA dehydrogenase activity (DBP). Two thiolases complete the pathway: the sterol carrier protein X (SCPX) reacts with multiple substrates, whereas acetyl-CoA acyltransferase 1 (ACAA1) only processes 3-keto-acyl-CoA-esters (Wanders et al. 2010).

Further important metabolic pathways in peroxisomes are the amino acid metabolism (Figure S1, green) and the purine metabolism (Figure S1, pink) which are, together with the fatty acid oxidation enzyme ACOX, high producers of cell toxic hydrogen peroxide (Figure S1, grey arrows; enyzmes: DAO, HAO, PAOX, PIPOX, XDH). The **antioxidant system** (Figure S1, violet) within peroxisomes antagonizes this toxic threat by oxidoreductases converting the H_2O_2 into oxygen and water. The most prominent enzyme of the antioxidant system is the catalase (EC 1.11.1.6) which was often used as a key marker for the presence of the peroxisome organelle. Furthermore, the antioxidant activity was the first function described in peroxisomes and is responsible for the name of this organelle (Duve, Baudhuin 1966; Schrader, Fahimi 2008).

1.6 Background, Relevance and Aims of This Thesis

1.6.1 Background and Relevance

Investigating the biology of the two photoautotrophic algae *C. velia* and *V. brassicaformis*, which are closely related to the predominating global human pathogens represented by the apicomplexan genus *Plasmodium*, should lead to a better understanding of its parasitic biology. Discrepancies can give information about putative drug targets in order to control, suppress or finally erase the parasite. In 2015, malaria still caused the death of a child every two minutes and 438,000 deaths worldwide with 90% in the hotspot Sub-Sahara Africa (WHO 2016). Fundamental milestones have been achieved in fighting malaria, thus there was a decline of malaria cases by 18% and a decline of malaria deaths by even 48% in 2015. In addition, there was a record of zero indigenous malaria cases in the European Region for the first time (WHO 2016). This success is linked with the introduction of effective prevention methods and pharmaceuticals against these parasites. For instance, artemisinin-based combination therapies (ACT) were assigned to 21% averted Malaria cases/deaths (WHO 2016). Artemisinin is an antimalarial sesquiterpene lactone peroxide extracted from the Chinese plant *Artemisia annua*, which was already applied to fever patients in China since 2000 years. This ancient knowledge described in historic Chinese medical documents formed the basis to the discovery of this substance in 1970s by Tu Youyou, whose work was therefore rewarded with the Nobel Prize for Medicine in 2015 (Hsu 2006; Tu 2011).

The most casualties of deaths caused by malaria are in children under five years. On this account, the vaccine RTS,S ("Mosquirix") was licensed for a pilot implementation in infants in October 2015 by the WHO. For a vaccination, the parasite must be attacked in an early stage of infection by the immune response. The recombinant protein-based vaccine Mosquirix was designed to target predominantly the parasitic circumsporozoite (CS) protein, which is found on the surface on *Plasmodium* sporozoites of the pre-erythrocytic phase during the invasion of hepatocytes. Thus, the vaccine prevents the infection of the liver and the outbreak of the clinical disease (Wilby et al. 2012).

Both examples (artemisinin, Mosquirix) demonstrate that the malaria disease never lose its topicality. Unfortunately, parasites often develop resistances against drugs after a period of time that turns the work of researchers into a race against time. Accordingly, there exist already reports about resistances against the artemisinin-containing medicament in five countries of the Greater Mekong subregion (Ashley et al. 2014; WHO 2016).

Previous drugs concentrated mostly on organelle functions such as the use of the antibiotic fosmidomycin which inhibits the MEP pathway of the plastid (Jomaa et al. 1999; chapter 1.5.2). Thereby, most reports highlighted the control of the malaria disease. Drugs targeting the apicoplast and possible resistances were recently reviewed in a study by Avinaba Mukherjee and Gobinda Chandra Sadhukhan in 2016. In this publication the authors conclude that *"[…] the development of new anti-malarial drugs targeting the apicoplasts may represent a significant breakthrough in drug development […]"*.

Inhibitions of mitochondrial functions were intensively investigated (Souza et al. 2009). For instance, the previously mentioned agent artemisinin attacks indirectly the alternative complex I NADH-quinone oxidoreductase of the electron transport chain in mitochondria by generating reactive oxygen species (Li et al. 2005). Further mitochondria-targeting drugs are the actually antifungal antibiotic atpenins, which are blocking the electron transfer in complex II (Miyadera et al. 2003). Furthermore atovaquone has an effect on the proliferation of protist parasites by collapsing the mitochondrial membrane potential (Fry, Pudney 1992; Srivastava et al. 1997; Kaneshiro et al. 2000; Siregar et al. 2015). Alternative oxidases in mitochondria of apicomplexan parasites can be effectively inhibited by the agents salicylhydroxamic acid, propyl gallate and 8-hydroxyquinoline (Roberts et al. 2004).

These examples show that the understanding of organelle function and evolution in apicomplexan parasites and their close relatives is of particular importance and paves the way for a targeted development of novel drugs.

1.6.2 Aims of this Thesis

My PhD thesis is focused on the evolution of three pivotal organelles (mitochondria, plastid and peroxisome) in 'connecting links' of apicomplexan parasites and dinoflagellates (Myzozoa). Therefore a **deep sequencing approach** of the genomes and transcriptomes was performed extracted from the photosynthetic apicomplexan algae *C. velia* and *V. brassicaformis* as well as from the non-photosynthetic dinoflagellate parasite *P. olseni*. Furthermore, the transcriptome of the peridinin-containing dinoflagellate *Prorocentrum minimum* was sequenced as a dinoflagellate reference organism. The sequencing of dinoflagellate nuclear genomes is bothered by their very large sizes of more than 100 Gb and high expanded gene contents (Hackett et al. 2004a; LaJeunesse et al. 2005; Shoguchi et al. 2015). However, parasitic organisms such as the genus *Perkinsus* have a reduced genetic content owing to their heterotrophic life style, which facilitates first sequencing of dinoflagellate genomes. The most prominent *Perkinsus* species *P. marinus* was already in the

spotlight of former studies (e.g. Matsuzaki et al. 2008; Joseph et al. 2010) and its draft-genome was sequenced with an eight-fold coverage (TIGR database, NCBI GenBank Accession AAXJ00000000.1), which allowed to identify the majority of genes from the plastid MEP pathway (Grauvogel et al. 2007). A high quality genomic dataset of the *Perkinsus olseni* was accomplished with a 100-fold coverage in order to get a reliable insight into their evolution and the **putative presence of a plastid** in this non-photosynthetic dinoflagellate. The presence of a plastid in this dinoflagellate genus is still under debate although numerous indicators for its existence are present such as the discovery of all genes of the plastid-specific MEP pathway in expressed sequence data (Grauvogel et al. 2007; Matsuzaki et al. 2008; Fernandez Robledo et al. 2011). Additionally to the *in silico* data analysis, the investigation of a putative plastid structure in *P. olseni* was conducted by laboratory experiments.

The apicoplast is a phylum-specific designation for a non-photosynthetic plastid in Apicomplexa. The discovery of the two photosynthetic algae *C. velia* and *V. brassicaformis* that are closely related to these parasites challenges the definition of the apicoplast. Their plastomes in combination with the species tree and phylogenies of plastid key markers should shed light on **plastid evolution** in this alveolate phylum.

Little is known about the composition of **mitochondrial genomes** in *C. velia*, *V. brassicaformis* and *P. olseni*. Due to the observation that the plastid genomes of both chromerids incorporate much more genes than plastid genomes of the closely related apicomplexan parasites (Janouskovec et al. 2010), a comparable finding for their mitochondrial genomes was expected. In the case of the mitochondrion of the oyster parasite *P. olseni*, a reduced gene content was assumed comparable to those of the early-branching parasitic dinoflagellates *Oxyrrhis marina* or *Hematodinium* sp. containing just three oxidative phosphorylation genes (*coxI*, *cox3*, *cob*). Therefore, all three sequenced genomes were tested by a combination of both *in silico* data analyses and laboratory experiments.

Finally, the comprehensive investigations about **the presence of peroxisomes** in all genome-sequenced alveolate phyla including the ciliates represent the first exhaustive *in silico* analysis of this neglected organelle in the eukaryotic superensemble.

2. Material and Methods

The following flow chart (Figure 8) represents the general workflow implemented in this study:

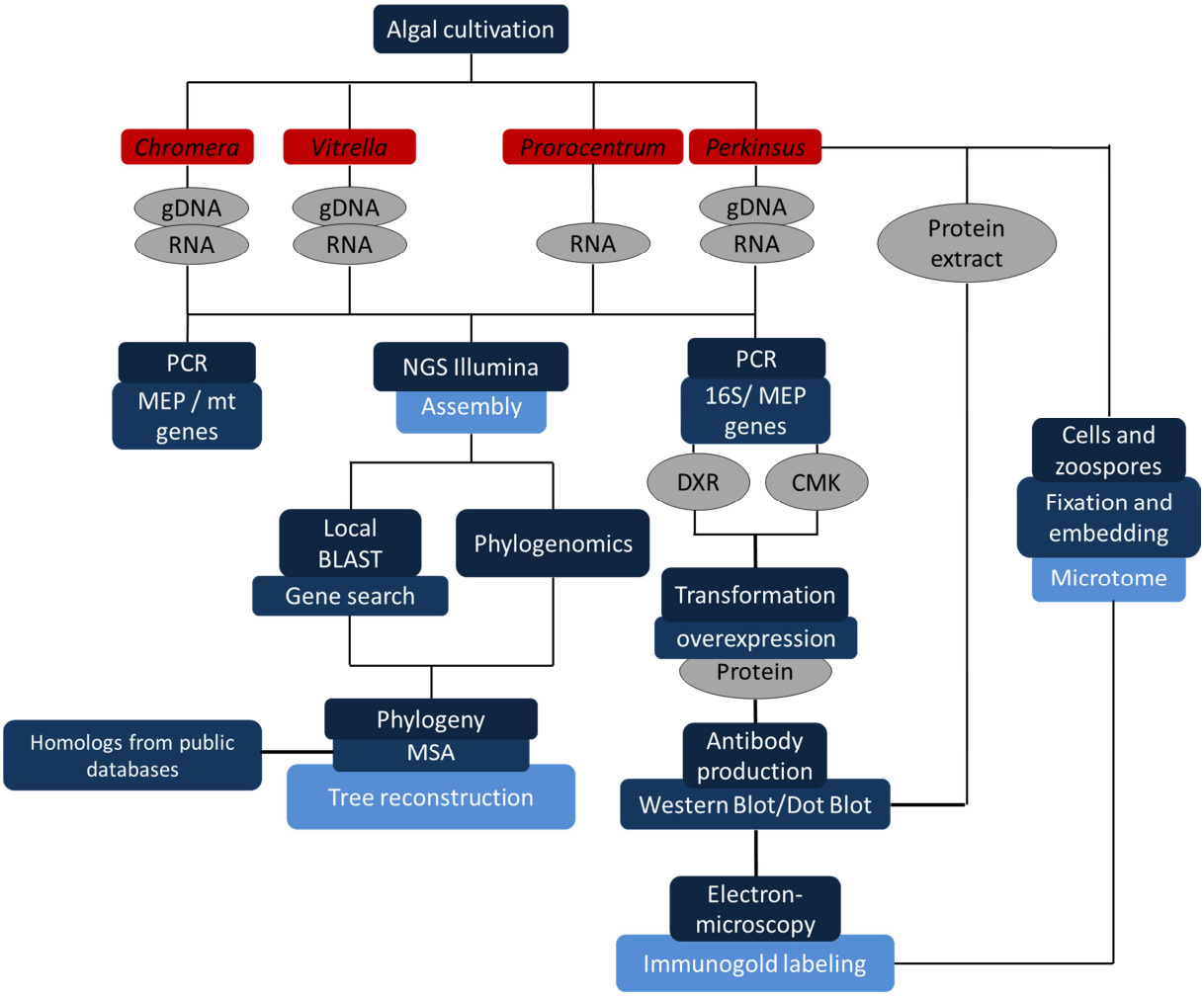


Figure 8: Flowchart of the working steps of this thesis. Abbreviations: gDNA, genomic DNA; MEP, non-mevalonate pathway; NGS, next generation sequencing; MSA, multiple sequence alignment. Further abbreviations are in compliance with the index and text.

2.1 Establishment of High Quality Genomes and Transcriptomes of Algal Key Species

2.1.1 Algal Cultivation and DNA/RNA Isolation

Algal Cultivation

In our group we cultivated the following protist strains: *Chromera velia* CCAP 1602/1, *Vitrella brassicaformis* CCMP 3155, *Prorocentrum minimum* CCMP 1329, *Perkinsus olseni* ATCC® PRA-181™. The marine protist strains were chosen from the respective cell collections (CCAP = Scottish Culture Collection of Algae and Protozoa; CCMP = Provasoli-Guillard National Center for Culture of Marine Phytoplankton; ATCC = American Type Culture Collection). The strain of the chromerid *C. velia* is an isolate from the Scleractinian coral *Plesiastrea versipora* of the Sydney Harbor, New South Wales, Australia. The second chromerid strain *V. brassicaformis* was likewise isolated from a coral animal of the Great Barrier Reef site. The collection site of the axenic dinoflagellate reference *P. minimum* is the Great South Bay, Long Island, USA. A clonal lineage strain of *P. olseni* was kindly provided by Christopher F. Dungan. The heterotrophic protist was isolated from a marine site of the Mie prefacture in Japan.

The cryo-conserved algal strains were recultivated in their appropriate growth medium. Table 1 shows a scheme of growth conditions for each organism.

Table 1: Summary of the culture conditions for four different protist strains: the photoautotrophic algal species *C. velia*, *V. brassicaformis* and *P. minimum* as well as the heterotrophic dinoflagellate *P. olseni*. The conditions of both chromerids are merged to one row.

Organism	Growth Medium	Temperature	Agitation	Light
Chromerida	L1 medium	22°C	✓	✓
***Perkinsus olseni* (axenic)**	DME/F12-3ps	22°C	x	x
***Prorocentrum minimum* (axenic)**	L1-Si medium	22°C	x	✓

Both chromerids were cultivated in 400 mL L1 medium in a 1l Erlenmeyer flask at 22°C and were shaken in a New Brunswick Scientific Innova 42 Incubator shaker at 100 rpm under continuous light. The dinoflagellate parasite *P. olseni* was cultivated in 5 to 50 mL 850mOsm/kg (29ppt) DME:Ham's F-12 *Perkinsus* sp. propagation medium containing 3% (v/v) fetal bovine serum (FBS) and 100 U µg/mL penicillin-streptomycin antibiotics for axenic growth in cell culture flasks at 22°C without shaking and light (Burreson et al. 2005). The axenic dinoflagellate strain *P. minimum* CCMP1329 was cultivated in L1-Si medium in

light without shaking. The cell material of the different algal strains was harvested to pellets via centrifugation and was subsequently stored in liquid nitrogen.

DNA and RNA Isolation

The cell material of the different algal strains was pestled in liquid nitrogen. The purification of the genomic DNA was performed with the DNeasy® Plant Mini Kit (Qiagen) and the isolation of the RNA was accomplished with the TRIzol® reagent (Life Technologies) following the manufacturer's instructions. The messenger RNA (mRNA) was isolated with the *PolyATract mRNA Isolation System III* (Promega).

2.1.2 Illumina Next Generation Sequencing and Sequence Assembly Techniques

The establishment of DNA and RNA libraries as well as the procedure of sequencing performed on different Illumina platforms was accomplished. The DNA libraries with a size of 450 bp were produced for the myzozoan strains of *C. velia*, *V. brassicaformis* and *P.olseni* according to the Illumina manufacturer's instructions (*Preparing Samples for Paired-End-Sequencing*). The DNA was fragmented using the Covaris S2 system and afterwards the fragments were ligated to specific adapters to diminish cross contaminations with previous Illumina runs. These constructs were purified and size selected on a gel before they were transferred to the cluster generation platform. The fragments were hybridized onto the flow cell by the Illumina Cluster Station. The Illumina paired-end sequencing of 110–150 bp was conducted on both Genome Analyzer IIx (GA) and Miseq sequencers (Table 2; Petersen et al. 2014).

The RNA libraries of 300 bp were established for *C. velia*, *P. olseni* and *P. minimum* due to the *TrueSeq RNA Sample Prep Guide* of Illumina. Prior to the first-strand cDNA synthesis the mRNA was fragmented by the Covaris S2 system. The fragments were tagged with specific adapters, quality controlled by the Bioanalyzer/Qubit and transferred to the cluster generation platform. The constructs were hybridized onto the flow cell by the Illumina Cluster Station and subsequently paired-end sequenced (100–150 bp) on GA, Miseq and Hiseq 2000 sequencers (Table 2; Petersen et al. 2014). The RNA libraries of 150–350 bp of *V. brassicaformis* were established by the external Vertis Biotechnology AG (Freising, Germany). The company performed the isolation of the mRNA, the normalization of the mRNA as well as the reverse transcription to cDNA before the construction of the libraries. These normalized libraries were used for paired-end sequencing (100–150 bp) on GA, Miseq and Hiseq 2000 sequencers (Table 2).

The Illumina sequence reads were converted into the FASTQ format and they were in general *de novo* assembled with VELVET 1.2.07 (Zerbino, Birney 2008). The sequence data were controlled for quality features (adapter and quality trimming) with the fastq-mcf tool of ea-utils (Aronesty 2011). Different assembly techniques were tested via alternating quality controls via the presence of specific markers in full length (e.g. MEP pathway). In some cases different runs were coassembled with VELVET like the final genomic datasets of *C. velia* (J2) and *P. olseni* (A3). In the final transcriptomic dataset of *C. velia* (E7), *P. olseni* (G2) and *P. minimum* (F5) the assembled contigs were extended and partly scaffolded by the pipeline of the program SSPACE 2.0 (Boetzer et al. 2011). In *V. brassicaformis* the quality of the genomic *de novo* assembly (JP06) could not be increased by any other assembly program, whereas the transcriptomic assemblies by VELVET and SSPACE were merged by MIRA 3.4.1.1 (Chevreux et al. 2004). Table 2 summarizes the different techniques used for every generated final dataset.

Table 2: Overview of the Illumina Next Generation Sequencing Project: genomic and transcriptomic data libraries and assemblies of four algal key species. The sequencing runs were accomplished on different machines due to updates of the techniques at the HZI. Assembly method choice corresponded to quality of full length target genes. The final assemblies were highlighted in bold. Abbreviations: PE = paired-end library, GA = Genome Analyzer.

Organism	Material	Library	Illumina run	Assembly method
Chromera velia	Genome	JP01	0.5 x 150 bp PE GA	VELVET 1.2.07
			1.0 x 110 bp PE GA	Coassembly (**J2**)
Chromera velia	Transcriptome	JP03	0.5 x 150 bp PE GA	VELVET 1.2.07
			0.5 x 110 bp PE GA	SSPACE
			0.5 x 100 bp PE Hiseq	SSPACE (scaffolding, **E7**)
Vitrella brassicaformis	Genome	**JP06**	1.0 x 150 bp PE Miseq	**VELVET 1.2.07**
			1.0 x 110 bp PE GA	Coassembly
Vitrella brassicaformis	Transcriptome		1.0 x 150 bp PE Miseq	VELVET 1.2.07
		JP05	1.0 x 110 bp PE GA	SSPACE
			1.0 x 150 bp PE Miseq	
		JP14	0.5 x 100 bp PE Hiseq	MIRA (padded, **D15**)
Perkinsus olseni	Genome	JP08	0.5 x 150 bp PE Miseq	VELVET 1.2.07
			1.0 x 110 bp PE GA	Coassembly (**A3**)
			2.0 x 110 bp PE GA	
Perkinsus olseni	Transcriptome	JP09	1.0 x 150 bp PE Miseq	VELVET 1.2.07
			0.5 x 100 bp PE Hiseq	SSPACE (**G2**)
Prorocentrum minimum	Transcriptome		0.5 x 150 bp PE GA	VELVET 1.2.07
		JP04	1.0 x 110 bp PE GA	Coassembly
			1.0 x 150 bp PE Miseq	
			(J. Tomasch, HZI)	
		JP13	0.5 x 100 bp PE Hiseq	SSPACE (scaffolding, **F5**)

2.1.3 Quality Control and Annotation of the Illumina Transcriptomes

Statistical values to validate the quality of the transcriptomal datasets were partly calculated by our cooperation partner Michael Jarek (coverage information) and another part (maximal contig length, N_{50} value, GC content) was evaluated by the web-based program Prinseq (Schmieder, Edwards 2011). Preliminary annotations of each sequence in the transcriptomic datasets were accomplished by the annotation platform of KEGG (KAAS, Moriya et al. 2007) and additionally for the transcriptome of *P. olseni* by the annotation pipeline of Blast2GO (Conesa et al. 2005).

2.1.4 Identification of Specific Genes in the Newly Established Transcriptomic and Genomic Datasets

We installed an application for local blast analyses (blast 2.4.0+; Camacho 2016; NCBI) in order to identify genes/proteins of interest. Appropriate query sequences from closely related species to the selected study organism were used to conduct tblastn (metabolic proteins) and blastn (16S rRNA, transcriptomic vs. genomic data) searches. Positive BLAST results were counterchecked by the blastx tool in the NCBI Web Server to validate the correctness of the results. The coding region of the respective scaffold/contig/node was separately saved and translated in the corresponding reading frame to obtain the protein sequence. These data were further used for multiple protein alignments, sequence analyses, phylogenetic analyses and primer design.

In the special case of the 16S rRNA as a key marker for the presence of a plastid, we performed a search with a 16S rRNA training set of the RDP classifier via a naïve Bayesian classification approach (Wang et al. 2007) in the genomic dataset of *P. olseni*. Furthermore, we browsed the unused reads of the Illumina genomic assembly of *P. olseni* for indications of 16S rRNA by means of the program SortMeRNA (Kopylova et al. 2012) and also again with the RDP classifier. Possible hits were counterchecked with the blastn tool in NCBI.

2.2 Sequence Analysis of Identified Genes and Proteins

2.2.1 Intron/Exon Annotations

Sequences of the same gene were compared on the transcriptomic and genomic level in order to detect potential spliceosomal introns. The intron borders were recognized by N-terminal ‘GT’ signal dinucleotides and by C-terminal ‘AG’ signal dinucleotides via a manual visual

inspection of the transcriptomic/genomic sequence comparison operated in the sequence analysis program MEGA 5 (Tamura et al. 2011). Introns can be classified into three phases depending on their location in the reading frame of the mature protein sequence: (1) location between two successive codons (phase 0); (2) location between the first and the second nucleotide of the codon (phase 1); (3) location between the second and the third nucleotide of the codon (phase 2; reviewed in Rogozin et al. 2012). In genes and proteins of interest introns were assigned to such a classification by means of the three-frame translation function in the restriction map tool of BioEdit v. 7.2.5 (Hall 1999). Comparing the mature protein sequence with the three-frame translation of the coding nucleotide sequence, it was possible to distinguish the different intron phases. The analysis of intron distribution and conservation in different species was accomplished by protein sequence alignments calculated by the MUSCLE algorithm in MEGA 5 (Tamura et al. 2011). The amino acid positions with introns were marked in the alignments and the corresponding intron phase was recorded.

2.2.2 Annotation of Signaling and Targeting Presequences

The localization of the encoded proteins was predicted by web-based programs examining the N-terminus of the translated sequences. In general, the protein sequences were loaded into SignalP v. 3.0 (Nielsen et al. 1997; Nielsen, Krogh 1998; Bendtsen et al. 2004) in order to confirm the presence of a presequence. Therefore, the program truncated the respective sequence to 70 aa and predicted the cleavage site of this presequence via the Hidden Markov Model technique (HMM). This program provides no information about the localization of the respective protein. This was done by the web-based prediction program TargetP v. 1.1 (Nielsen et al. 1997; Emanuelsson et al. 2000), which also provided the information about a potential cleavage site of the targeting sequence (TPlen = predicted presequence length). Four different scores were calculated for four different options of localization: cTP (chloroplast transit protein), mTP (mitochondrial transit peptide), SP (signal peptide containing protein of the secretory pathway) and other (any other localization). The highest score was assigned to the most probable localization. However, the results were cross-validated by the internal reliability class (the relationship between the scores; 1-5 with 1 for the strongest prediction) in order to check the reliability of the predictions. Frequently, other prediction programs were consulted in parallel in challenging cases: iPSORT (Bannai et al. 2002) and PSORTII (Nakai, Horton 1999). Both additional programs were also very helpful in the determination of N-terminal and C-terminal peroxisomal targeting sequences.

In dinoflagellates bipartite targeting sequences can be classified into three classes dependent on the number of hydrophobic transmembrane domains within the pre-sequence (Patron et al. 2005). These domains were located with the program SOSUI (Hirokawa et al. 1998) and Kyte-Doolittle hydropathy plots of the program ProtScale in ExPASy (Gasteiger et al. 2005) as well as the program pepwindowall in EMBOSS (Rice et al. 2000). In addition to a targeting sequence in dinoflagellate sequences the presence of a spliced leader (SL) sequence as a signal for nuclear-encoded genes was identified via the search function in MEGA 5 (Tamura et al. 2011) with the appropriate query SL sequences of *P. marinus* (Zhang et al. 2011) or other dinoflagellates (Zhang et al. 2007).

2.2.3 Venn Diagram Analysis

The Venn diagram of primary and secondary endosymbiotic derived plastomes in eight different species (*C. paradoxa* [NC_001675]; *C. crispus* [NC_020795]; *A. thaliana* [AP000423]; *G. theta* [NC_000926]; *E. huxleyi* [AY741371]; *T. pseudonana* [EF067921], *P. falciparum* [Wilson et al. 1996; Arisue et al. 2012]; Dinoflagellate minicircles [Hackett et al. 2004]) was generated via a self-written R-script (Figure S2). R is a software package implementing statistical methods. Therefore, numerous free subprograms and packages are available online. In this case the subprogram 'venneuler' was used in order to generate a comprehensive and proportional Venn diagram for large datasets (Wilkinson 2012). In the first instance, a presence/absence table of plastome-encoded genes was generated and imported in .csv format into the R command line. The list was transformed into a logical matrix for the presence (= TRUE) and the absence (= FALSE) of a plastome gene. This matrix served as the input into the venneuler function to generate the Venn diagram. The proportions of the graphical output were used as a model for the final self-made illustration finished in Microsoft PowerPoint 2010. With the same techniques reduced datasets and their consequential Venn diagrams were reconstructed.

2.2.4 Amplification of Genetic Material via PCR Techniques

Conventional Polymerase Chain Reaction (PCR)

Genetic material and genes of interest were amplified using the *Crimson Taq DNA Polymerase Protocol* (New England Biolabs GmbH, NEB). Mastermix and thermocycler program are summarized in the Figure 9.

The amount of applied DNA is dependent on the concentration of the template and is one adjusting factor in the mastermix. The PCR procedure starts with the initial denaturation of the template DNA in order to make it accessible to the primer pair. A repetitive cycle with three steps forms the core of the PCR reaction. The steps are composed of the denaturation of the template DNA (95°C), the annealing of the primer pairs (45 – 68°C) and the extension of the complementary strand starting from the annealed primer pairs in 5' to 3' direction (68°C). In general these steps are repeated for 30 cycles, but for the detection of low-copy targets, the cycle number can be elevated up to 45 cycles. With these cycles the targets can be amplified in an exponential manner with up to 10^8 copies per product. The final extension step (68°C, 5 minutes) terminates the PCR reaction. The Crimson Taq DNA Polymerase has a terminal transferase activity and generates A-overhangs at the ends of the targets, which can be useful for further cloning experiments. Designed Primers and their corresponding melting temperatures for targets of interest are assembled in Table S1.

Mastermix

Component	25 µL reaction	Final concentration
5x Crimson Taq reaction buffer	5 µL	1x
10 mM dNTPs	0.5 µL	200 µM
10 µM Forward Primer	0.5 µL	0.2 µM
10 µM Reverse Primer	0.5 µL	0.2 µM
Template DNA	1 ng – 1 µg	<1,000 ng
Crimson Taq DNA polymerase	0.125 µL	1.25 units/50 µL PCR
Nuclease-free water	to 25 µL	

Thermocycle Conditions

Initial denaturation		95°C	30 seconds
30 cycles		95°C	15 – 30 seconds
		45 – 68°C	16 – 60 seconds
		68°C	60 seconds/kb
Final extension		68°C	5 minutes
Hold		4 – 10°C	

Figure 9: Composition of PCR master mix and thermocycle conditions according to the Crimson Taq DNA Polymerase Protocol (NEB).

Inverse PCR Experiments for Circular Products (iPCR)

Based on the Illumina mitochondrial coding sequences the possible circularity of mitochondrial molecules was tested for *C. velia*, *V. brassicaformis* and *P. olseni* via the inverse PCR technique. Therefore primer pairs were developed, which are not directed to each other but to opposite directions beginning in the known coding regions of mitochondrial protein coding genes (*coxI*, *cob*; Table S2). For the inverse PCR experiments again the *Crimson Taq DNA Polymerase Protocol* (New England Biolabs GmbH, NEB) was used. In the case of a single circular DNA template, one band should be obtained if the unknown region is of an amplificable size.

Tailing PCR Experiments for Linear Products

Reciprocally, the possible linearity of the mitochondrial molecule of *C. velia* was tested. Therefore, the hypothetical ends of the linear molecule were extended using only one specific 2´-desoxyribonucleosid-5´-triphosphate (dNTP) by the activity of the enzyme *Terminal Deoxynucleotidyl Transferase* (TdT). This enzyme extends the ends of a linear molecule multiple times, hence it forms a signal sequence of repetitive dNTPs. The tailing reaction was accomplished using the *Terminal Deoxynucleotidyl Transferase, Recombinant Protocol* (Promega), 2 pmol of template DNA as well as two approaches of tailing experiments with either dATPs or dCTPs. The mixture was incubated with 10 – 20 units of the TdT enzyme for 60 min at 37°C. The terminal extension step occurs at 70°C for 10 min and the reaction product was stored on ice or in the fridge at 4°C. Specific primers were established to enable a nested PCR approach (*Crimson Taq DNA Polymerase Protocol*) for the amplification of the molecule's ends (Table S3) which are composed of an anchor sequence linked to multiple dATPs/dCTPs for the first primer and simply the anchor sequence for the second primer. Positive PCR results verified by agarose gel electrophoresis would indicate the linearity of a molecule.

cDNA Synthesis via Reverse Transcriptase Activity (RT-PCR)

First-strand cDNA (complementary DNA) was synthesized out of total RNA with the protocol of the *SuperScript™ III First-Strand Synthesis System for RT-PCR* (Invitrogen). The in the Kit included RNA-dependent DNA Polymerase uses an oligo(dT) primer, which binds to the complementary polyA-tail of the eukaryotic mRNA. The primer/RNA/dNTP mixture is heated at 65°C in RNAse-free water and afterwards placed on ice to avoid secondary structures and to ensure easily accessible single-stranded conditions. Afterwards, the cDNA synthesis mix is added and the primers are annealed at 50°C to the RNA fragments. In the next step, the reverse transcriptase connects to the primer sites and starts the production of the complementary strand. The original mRNA is subsequently digested by the enzymatic reagent RNaseOUT™ of the Kit. The final product of the reaction is a double-stranded molecule, whose sequence information can be directly translated to the mature protein information. The reaction is terminated at 85°C and afterwards briefly chilled on ice. In order to purify the cDNA, the enzyme RNase H is added to the final product in order to eliminate remnant RNA fragments (37°C, 20 min). cDNA of *C. velia* was prepared in order to complete the sequence information of partial MEP proteins in the Illumina transcriptomic dataset.

2.2.5 Processing of PCR Products

Separation of Genetic Material and PCR Products via Agarose Gel Electrophoresis

Most of the genetic material was separated on 1–1.5% agarose gels containing the DNA intercalating dye ethidium bromide that can be visiualized under UV light. The PCR products are pH stabilized by a loading buffer, which also serves as an internal colored running time control by building a solvent front of anionic dyes. The size of the products can e.g. be estimated via a loaded 1 kb ladder (Invitrogen) size standard. In general, the gel was run at 120 V time-adjusted corresponding to the expected size of the product.

Purification of PCR Results

The PCR products were purified with the column-based *QIAquick PCR Purification Kit* (Qiagen). This Kit includes also all chemicals to elute specific PCR bands out of the agarose gel if there appear multiple product bands. The Kit purifies the PCR products in three simple steps: (1) binding the DNA on a silica membrane under high-salt conditions provided by the binding buffer; (2) washing the PCR product from impurities such as primers via a washing buffer; (3) finally eluting the pure DNA from the silica membrane under low-salt conditions. The same principle is applied for gel extracted PCR products.

Cloning of PCR Products

The A-overhang of PCR products generated by the Crimson Taq DNA Polymerase was used in combination with the TOPO® TA cloning® Kit (Invitrogen) to insert the product into the pCR™2.1-TOPO® vector. The ligation of the insert into the vector was accomplished by the T_4 ligase overnight at 4°C. Vector and insert were transformed into NEB® Turbo Competent *E.coli* via the heat shock method and subsequently incubated overnight at 37 °C. The heat shock method was performed at 42°C for 30 seconds. These transformed bacterial cells were inoculated in lysogeny broth medium to initiate the growth of the cells at 37 °C. Afterwards, the cells were plated in different concentrations on LB medium plates containing the broad-spectrum antibiotic Ampicillin for which the NEB® Turbo cells should have a resistance on the transformed vector. This system is also suitable for blue/white screening by the α-complementation of the β-galactosidase gene using the pCR™2.1 vector. In this case the transformed cells were inoculated onto AIX plates containing LB + Ampicillin + Isopropyl-β-D-1-thiogalactopyranoside (IPTG) + the organic compound X-gal. Successful ligated inserts disrupt the α-complementation process and result in white colonies, whereas vectors without

insert form blue colonies, because the complemented β-galactosidase enzyme can cleave the colorless X-gal into its blue-colored product.

Colonies were picked and cultivated in liquid LB+Amp test tubes at 37 °C overnight. In the next step the transformed plasmids are regained by a plasmid preparation method accomplished by the use of the NucleoSpin® Plasmid Kit (Macherey-Nagel). Therefore, the bacterial cells were broke down via SDS/alkaline lysis and subsequently centrifugated in order to get a supernatant containing the plasmids. The lysate was loaded on a silica membrane spin column and the DNA was bound under high-salt conditions. After washing steps, the vectors were eluted with a slightly alkaline buffer.

In order to check the presence of the favored insert in the vector, the plasmids were digested with the EcoRI restriction enzyme. The palindromic cleaving sequence of EcoRI ($\frac{G:AATTC}{CTTAA:G}$) flanks the insert including only a few partial sequences of the original vector. The digestion reaction occurs for approximately an hour at 37 °C and the isolated inserts can be subsequently visualized via gel electrophoresis. On the basis of their expected size, the fragments of interest were sequenced with the Sanger method by SEQLAB Sequence Laboratories Göttingen GmbH.

In this study, PCR products were in general cloned according to this method in order to increase the availability of the product and to get pure sequences.

2.3 Immunocytochemical Studies

Our intention was to get a clear visual evidence for the presence of a plastid structure in *Perkinsus* sp.. Therefore antibodies were generated against two proteins of the plastid-specific MEP pathway (DXR and CMK) to use them in immunogold labeled electron microscopic experiments.

2.3.1 Overexpression of the MEP Proteins DXR and CMK

Preparation of the Target Sequences

Cells of *Perkinsus marinus* (strain ATTC®50439™) were cultivated at the same conditions as the cells of *Perkinsus olseni* (ATCC® PRA-181™; chapter 2.1.1). Genomic DNA was extracted from these cells following the instruction of *DNeasy® Plant Mini Kit* (Qiagen). We generated cDNA from the isolated genomic DNA of *Perkinsus marinus* following the instructions of the *SuperScript™ III First-Strand Synthesis System for RT-PCR* (Invitrogen, chapter 2.2.4). Conventional PCR methods were applied to amplify the full length cDNA

sequences of the MEP proteins DXR and CMK. Primer pairs were designed (Table S1) in order to accomplish a cloning experiment into an appropriate overexpression vector, which is directed by an N-terminal BamHI restriction site and a C-terminal HindIII restriction site. Furthermore, the forward and reverse primers carried in both cases (DXR and CMK) a four nucleotide overlap for the restriction enzyme digestion. Because both sequences should target the mature protein sequence potentially located in a plastid structure, the primers did not include any bipartite targeting signal, which is found in mRNAs of these MEP genes. The PCR products of the DXR and CMK sequences were ligated into the pCR2.1 vector, transformed into competent *E. coli* cells and re-isolated by plasmid preparation and restriction enzyme digestion (chapter 0.02.2.5). The sequences were checked by SeqLab Sanger sequencing and a full length 'master clone' aligned to the known query sequences of *P. marinus* (Matsuzaki et al. 2008) with the least number of PCR errors was determined. The construct of the DXR and CMK sequence within the pCR2.1 vector was double digested with the restriction enzymes BamHI and HindIII (Thermo Fisher Scientific). The isolated cDNA sequence was separated via gel electrophoresis and extracted following the instructions of the *QIAquick PCR Purification Kit* (Qiagen; chapter 0.02.2.5).

Preparation of the Overexpression Vector

We selected the overexpression vector pQE80L (Qiagen) as the appropriate reaction partner for our cDNA sequences. In a two-step approach, the vector was digested with HindIII and afterwards with BamHI (Thermo Fisher Scientific). In both cases the digested vector DNA was purified following the instructions of the *QIAquick PCR Purification Kit* (Qiagen; chapter 0.02.2.5). The linearized vector sequence was checked via agarose gel electrophoresis. After the digestion, there remain phosphate residues at the 5'-ends of the sequence. In order to prevent the self-ligation of the vector, because of these phosphate residues, the vector was dephosphorylated by the Shrimp Alkaline Phosphatase (SAP, Affymetrix) at 37 °C for one hour. The reaction was stopped by adding a mix of 10 µL 10% SDS, 2µL EDTA and 1 µL Proteinase K followed by another incubation step at 37 °C for 20 min. All enzymes of the reaction were inactivated at 68 °C for 15 min.

Test ligations were conducted for the dephosphorylated and the cut vector using the enzyme T_4-ligase. The incubation of this reaction was accomplished overnight at 4 °C. The result of the ligation experiments were separated via agarose gel electrophoresis (chapter 0.02.2.5). In the case of the cut vector, which underwent no dephosphorylation step multiple bands are

expected as vector-to-vector ligations can occur in multiple forms. The dephosphorylated vector should show an expected band with a size of approximately 4.7 kB.

Production of DXR/CMK + pQE80L Vector constructs

The BamHI/HindIII digested target sequences and the BamHI/HindIII digested and dephosphorylated vector were ligated with the T_4-ligase overnight at 4 °C. The next day the reaction was stopped at 75 °C for 10 min. The constructs were transformed into NEB® Turbo *E. coli* cells via the heat shock method and re-isolated by plasmid preparation (chapter 0.02.2.5). The test digestion with the restriction enzymes BamHI and HindIII with a subsequent separation on an agarose gel should confirm the success of this experiment.

Production of Competent Bacterial Cells for the Protein Overexpression Step

We selected the *E. coli* strain BL21 codon+, because it was designed to enhance the expression of eukaryotic proteins that contain codons rarely used in conventional *E. coli* strains. Essentially, these bacterial cells had to be made receptive for free DNA molecules. As a start, BL21 codon+ cells were cultivated overnight at 37 °C in sterile flasks with 20 mL LB medium and 20 µL of the broad-spectrum antibiotic chloramphenicol, for which this bacterial strain is drug-resistant. The next day, precultures of 50 mL LB and 50 µL chloramphenicol were inoculated with 1 mL of the overnight-culture and incubated at 37 °C for two hours. Within these two hours the optical density at a wave length of 600 nm (OD_{600}) of the preculture was regularly checked for the log-phase of bacterial growth. After achieving the log-phase, the whole volume of the preculture was transferred to a sterile centrifuge beaker and placed for 10 min on iced water. Reagents, which were used later in this approach, were also placed in ice water. Simultaneously, a part of the preculture was inoculated with an inoculation loop on a LB plate with chloramphenicol. The cells in the centrifuge beaker were centrifuged at 4,500 xg for 10 min at 4 °C to form a pellet. The supernatant was discarded, the pellet was scraped off with a silicon brush and resuspended in 25 mL of ice-cold 100 mM calcium chloride ($CaCl_2$) solution. The cells were centrifuged again at 4,000 xg for 10 min at 4°C. The supernatant was again discarded and the pellet was transferred and resuspended in 300 µL of 10% ice-cold glycerol. In 40 µL steps, the BL21 codon+ suspension was transferred into 1.5 mL Eppendorf vessels. The competent cells could be used directly or were stored at -80 °C with previous cryogenic freezing in liquid nitrogen.

Overexpression of DXR/CMK-pQE80L Constructs in Competent BL21 codon+ Cells

The constructs of DXR/CMK+pQE80L were transformed into the competent BL21 codon+ cells via the heat shock method (chapter 0.02.2.5). In order to select for both the chloramphenicol-resistant (Cam) BL21 codon+ cell lines and the ampicillin-resistant (Amp) pQE80L-vector, we cultivated the transformed bacterial cells overnight at 37 °C on LB-plates additionally containing these antibiotics. Precultures of 5 mL LB+Cam+Amp were inoculated with colonies grown on plates the day before. The cultures were again incubated overnight at 37 °C. 50 µL of the precultures were transferred into 10 mL LB+Cam+Amp of the main expression culture and incubated at 28 °C. The growth of the bacterial cells was monitored steadily by measuring the OD_{600} value with a photometer. At a OD_{600} value of approximately 0.1, 50 µM IPTG was added to the bacterial suspension and incubated at 25 °C for 22 hours shaking at 100 rpm in order to induce the expression of DXR and CMK. The cultures were centrifuged at 12,000 xg and 4 °C for 5 min in 2 mL Eppendorf vessels.

The expressed proteins were prepared following the instructions of the *QIAexpress® protein purification system* for pQE vectors (Qiagen). The proteins are linked at their N-terminus with a 6xHis-Tag provided by the pQE80L-vector. Such tagged proteins have a high affinity and specifity to nickel-nitrilotriacetic acid (Ni-NTA) spin columns. First, the cells in the pellet were broken down by the lysis buffer of the Kit. The linked proteins in the lysate were bound under native conditions onto the Ni-NTA spin columns, washed and eluted with the buffer system provided by the Kit. Samples of the flow through from the lysate, the first and second washing step, as well as the first and second elution step were mixed with Lämmli loading buffer + β-mercaptoethanol and loaded on a SDS polyacrylamide gel for electrophoresis (SDS-PAGE; 12% stacking/separation gel; apparatus: Mini-PROTEAN Tetra Cell, Bio-Rad Laboratories). With this technique proteins are separated according to their size in an electrical field (stacking gel: 100 V, 10 min; separation gel: 120 V, 60 min) compared to a loaded size marker PAGERuler® unstained (Fermentas). For the protein staining the gel was transferred into the first fixative/decolorizer I (10% glacial acetic acid, 40% ethanol) for ~ 20 min. The fixative was discarded and the protein bands were dyed with *Coomassie® Brilliant Blue R250* (Serva) solution containing 20% acetic acid. The staining was accomplished on a shaker for three hours. Afterwards the staining solution was discarded and the bright staining of the gel was reduced by washing with fixative/decolorizer I for 15 min. Furthermore the gel was additionally washed with fixative/decolorizer II (10% glacial acetic acid, 20% ethanol) for ~15–20 min. The fixatives/decolorizers were discarded after each step and the gel was finally washed two times with water. The blue stained protein bands were now clearly visible

and could be compared to the size standard. The DXR (~42/44 kDa) and CMK(~32/33 kDa) proteins were only sparsely recognizable via SDS-PAGE indicating that the proteins are accumulated in so-called inclusion bodies of the *E. coli* cells in the insoluble fraction (pellet). New SDS-PAGE approaches including soluble and insoluble fractions confirmed the presence of the protein in the insoluble fraction and thus in inclusion bodies. Therefore, the proteins were again purified with the Kit under denaturing conditions including buffers to convert the proteins in the inclusion bodies into a soluble form. The eluate was separated via SDS-PAGE showing only a weak protein concentration of both target proteins. The final protein concentrations of both protein eluates under native and denaturing conditions were determined with the Bradford reagent (Bio-Rad Laboratories). The optical density was measured at a wavelength of 595 nm.

2.3.2 External Overexpression and Antibody Production for DXR and CMK

In our laboratory, the overexpression of the two MEP proteins DXR and CMK resulted only in protein concentrations < 1 µg and we were not able to produce the protein on a larger scale. For this reason, we decided to assign the overexpression and antibody production job to the company GeneCust Europe (Luxembourg). Therefore, we provided a high amount of our antigen + pQE80L constructs by amplifying them in chemical competent NEB cells followed again by the preparation of the construct in order to receive 5 µg of the required target. The polyclonal antibody production was conducted in New Zealand rabbits, with two rabbits that were immunized per antigen. After nine weeks of immunization, the blood serum of each step and bleeding was collected and the quality was checked via ELISA and Western Blot analyses. We received two polyclonal antibodies per antigen and tested them internally against the produced DXR and CMK proteins via Western Blot analysis.

2.3.3 Dot Blot and Western Blot Analyses as Internal Antibody Quality Controls

The lyophilized polyclonal antibodies were resuspended in 5 mL 1xPBS and inverted multiple times. Five aliquots à 1 mL antibody solution were assembled and stored at -20 °C.

Dot Blot Analysis

Before starting the immunogold electron microscopic experiment, it had to be investigated if a specific fixative would possibly influence the antigen-antibody-binding reaction. We tested five different fixatives: (1) 3% formaldehyde; (2) 1% formaldehyde; (3) 1% glutaraldehyde; (4) 1% formaldehyde + 0.2% glutaraldehyde; (5) 1% formaldehyde + 0.5% glutaraldehyde. A stack of Whatman paper was soaked in blotting buffer (Tris-Glycin + 0.0375% SDS +

methanol) and placed free of air bubbles in a petri dish. A piece of pvdf membrane was activated with methanol and afterwards placed on the soaked Whatman paper stack. Per dot 2 µL of antigen (10 ng of the DXR protein) were dropped on the membrane. These dots were dried for one hour. Afterwards, the membrane was trimmed, thus each piece carried one protein dot. These dots were exposed to the different fixatives in smaller petri dishes for one hour in the fridge. Subsequently, the fixative was collected and the membranes were washed in 1xPBS + 10 mM Glycin for several minutes. The dots were blocked with 1% milk powder solution overnight in the fridge. The next day the six dilutions of one DXR antibody (RB3152) were prepared (1:50, 1:100, 1:1000, 1:2000, 1:5000, and 1:10000). The dilutions were dispensed in 36 x 2 mL Eppendorf vessels. The dot membranes of the previous day were transferred into the vessels and shaked for 90 min. The antibody dilutions were discarded and the membranes were washed three times with 1x PBS + Tween-20 (PBS-T) for 10 min. The second anti-rabbit IgG antibody was diluted 1:2000 in PBS-T and 1 mL was added to each Eppendorf vessel. The incubation with the second antibody was accomplished for one hour on a shaker and discarded afterwards. In the next step, the membranes were washed again three times for 10 min in PBS-T. The little membrane pieces were welded into an UV-permeable film adding 2 mL of the detection reagent *Lumi-Light Western Blotting substrate* (Roche). The second antibody carries the HRP enzyme, which generates chemiluminescence by activating an appropriate substrate. The chemiluminescence was documented by the Image Reader LAS-3000 (FujiFilm).

Western Blot Analysis

On the basis of this approach it is possible to test if antibodies generated against the *P. marinus* proteins would also target the two MEP proteins of our key species *P. olseni*. Therefore, the proteins from harvested cell pellets of both cultures were extracted and the binding behavior of the MEP antibodies against these protein extracts in comparison to the pure protein was tested.

To the protein extraction buffer (150 mM NaCl, 100 mM HEPES, 5 mM EDTA and 5% Glycerol) 0.1 % Tween-20 were added and mixed with ¼ tablet of the *cOmplete™, EDTA-free Protease Inhibitor Cocktail* (Roche). The cell pellets were pestled in liquid nitrogen and extraction buffer was added in a ratio of ~ 1:1. The mixture was incubated on ice for 15 min and continuously vortexed. Afterwards, the solution was centrifuged at maximum speed and 4 °C for 10 min.

The protein concentration of the obtained protein extract was determined using the Lowry Assay method following the instructions of the *Pierce™ Modified Lowry Protein Assay Kit* (Thermo Fisher Scientific). The different concentration values were measured against a standard and a calibration series of the bovine serum albumin (BSA) protein (3.3 µg, 10 µg, 30 µg and 50 µg) at a wavelength of 750 nm. The proteins and protein extracts were separated on a 12% SDS-PAGE gel with following Coomassie-staining and on a 14% SDS-PAGE gel for the Western Blot analysis. The MEP protein was loaded on the gel with a concentration of 500 pg, 5 ng and 50 ng. The extract of *P. marinus* was loaded with a concentration of 25 µg and 50 µg protein, whereas the concentration of the *P. olseni* extract started with 50 µg, followed by 80 µg and finally 100 µg protein. All samples were mixed with Lämmli buffer + β-mercaptoethanol and boiled up to 95/99 °C.

For the Coomassie-staining the 12% SDS gel, the same protocol as in chapter 2.3.1 was applied. The 14% SDS gel for the Western Blot analysis was transferred to the 1x blotting buffer (Tris-Glycin + methanol) and shaked for a few minutes. The *Roti®-Fluoro pvdf membrane* (Roth) was briefly activated in methanol and afterwards shaked in 1x blotting buffer together with Whatman Chromatography Paper (Roth) and the fiber pads for the blotting apparatus (Bio-Rad). For the blotting apparatus a gel sandwich was set up into the provided cassette in the following order: fiber pad, filter paper, gel, pre-wetted membrane, filter paper and finally again a fiber pad. The blotting process proceeded in 1x blotting buffer at 90V for 90 min in a cooled environment. The successful transmission of the proteins from the gel onto the membrane was checked with the help of the marker *PageRuler™ Plus Prestained Protein Ladder* (Thermo Fisher Scientific). The membrane was transferred to the blocking solution (skim milk powder [Sigma-Aldrich] in PBS-T) and blocked overnight in the fridge. The next day, the blocking solution was discarded and the membrane was wetted with PBS-T containing the first antibody against the protein of interest in a dilution of 1:1000. The blot was shaked under darkened conditions for 1-2 h. Afterwards, the antibody solution was discarded and the membrane was washed three times with PBS-T for 10 min. The second antibody (anti-rabbit IgG, HRP-linked, Cell Signaling) was diluted 1:2000 in PBS-T and added to the membrane. Again, the blot was shaken under darkened conditions for 1 h. The antibody solution was discarded and the membrane was washed once more three times with PBS-T for 10 min. The blot was welded into UV-permeable foil with the detection reagent *Lumi-Light Western Blotting Substrate* (Roche). The chemiluminescence was documented by the Image Reader LAS-3000 (FujiFilm).

2.3.4 Immunogold Electron Microscopy

Initially, the appropriate cell stage, the motile zoospores, had to be isolated from *P. olseni* cell cultures. In *Perkinsus* spp. the motile cell stage can be induced by the *Alternative Ray's Fluid Thioglycollate Medium* (ARFTM), which enhances the enlargement of the zoospore-containing hypnospore cell stages. Dilutions series of the cell culture were exposed to the ARFTM + lipid mixture + penicillin-streptomycin medium for ~ 48 hours (Dungan, Bushek 2015) and continuously checked for the amount of zoospores. The zoospore-enriched cell cultures were harvested and instantly fixed in the appropriate fixative (chapter 2.3.3). The fixed cells were dehydrated via ascending aceton series, were embedded in resin, polymerized and finally trimmed by an ultramicrotome (Ultratrim, Reichert) to ultra-thin sections. Each section was transferred to a coated nickel grid. On this grid the sections were incubated with the first polyclonal antibody against the antigen/protein of interest. Protein A which binds to the first antibody was linked to 15 nm colloidal gold particles. These gold particles have a high electron density and should appear in electron micrographs as high contrast dark spots in places, where the antigen is detected by the first antibody. The transmission electron micrographs (Zeiss EM910) were done at an acceleration voltage of 80 kV after a counter-staining of the probes with 4% aqueous uranyl acetate.

2.4 Bioinformatic Analyses

2.4.1 Phylogenomic Analyses

We performed phylogenomic analyses in order to reconstruct a species tree as a reference for the single gene phylogenies. Different data partitions were used to identify specific orthologous genes in a selection of appropriate organisms: mt-group (origin of α-proteobacteria), cp-group (origin of cyanobacteria), no outgroup-group (eukaryotic-specific genes), outgroup archae 1 (archaeal origin, single-copy genes), outgroup archae 2 (archaeal origin, multicopy gene, ancient duplications in eukaryotes) and ribosomal proteins. The phylogenomic tree was reconstructed as a concatenated tree in a supermatrix approach. The genomes (or transcriptomes) of different organisms were scanned for several sets (partitions) of orthologous genes. Every alignment to each selected gene was assembled with a specific taxon sampling, in this case alveolate species. A supermatrix (concatenation of single genes) was built by likelihood methods which consider evolutionary rate heterogeneity across the

single genes in this matrix. Non-existing positional information for one gene in taxa were coded as missing (Delsuc et al. 2005).

Subsequently, the tree reconstruction of this supermatrix was accomplished. All taxa of interest were selected in an OTU (Operational taxonomic unit) file. The program SCaFoS (Selection, concatenation and fusion of sequences, Roure et al. 2007) created a new aligned file, which only contained the OTU sequences. All single alignments were concatenated to one file to perform a first draft version of phylogenetic analysis. Preliminary test trees (PAUP, Treefinder) were generated by the scripts 'ali2paup' and 'ali2TF', respectively. The test trees were checked according to the taxon sampling, e.g. if some species were redundant and could be deleted or fused to chimera (same species, genus or family). The corrected taxon sampling was again used in SCaFoS, where the right taxa names and the concatenation of the files was accomplished. These files can be used by the MUST-package (Management utilities for sequences and trees, Philippe 1993). The 'gblocks' tool (Talavera, Castresana 2007) of this package eliminated all poorly aligned positions and divergent regions in each alignment. The now trimmed alignments were again used by the SCaFoS program to perform a concatenation of the datasets. For each trimmed alignment a bootstrapped RAxML-tree (ali2raxml) was built using a hundred bootstraps. These trees helped in the decision to define a default sequence for an OTU if there is the case like e.g. partial or paralogous sequences.

In the next step a congruence test was performed using PERL-scripts. In this test a bootstrapped reference tree of the concatenated alignment file was generated. Afterwards the script compared every single gene tree of the analysis with the concatenated tree thus highlighting incongruences supported by a bootstrap value of $\geq 70\%$. The incongruences were checked by determining if they are the result of tree building artefacts such as long branch attraction (LBA) or Nearest Neighbor Interchange (NNI) or if they rather represent a contamination or a HGT that need to be eliminated.

The species tree was finally calculated with the CAT-GTR model implemented in the program PhyloBayes. In contrast to other models, the CAT model does not assume, that all sites of a protein evolve under the same substitution process, but that different sites in this protein can be assigned to different substitution profiles in order to highlight site-specific features within protein evolution (classification into categories; Lartillot, Philippe 2004). Upstream of this model a dirichlet process is used in order to determine the total number of amino acid classes (e.g. hydrophobic, polar or positively charged) and the respective amino acid substitution profiles. The substitution rates from one amino acid to another were estimated by the GTR (Generalised time reversible) codon substitution model, which allows

variable instantaneous rates of substitution between each amino acid pair learnt from the given dataset (Le et al. 2008). This CAT-GTR profile mixture model fits statistically better to large datasets then empirical substitution matrices such as the WAG or LG model (Lartillot, Philippe 2004).

2.4.2 Phylogenetic Analyses

The alignments of the genes of interest were generated with ClustalW (Thompson et al. 1997) and manually refined using the ED option of the MUST package (Philippe 1993). The MUST-implemented g-blocks tool was used to eliminate highly ambiguous and variable sites in the alignments (Talavera, Castresana 2007). We performed maximum likelihood analyses within RAxML v7.9.5 (Stamatakis, Alachiotis 2010) in general under a LG+F+Γ4 model based on the LG-matrix of amino acid replacements (Le, Gascuel 2008) and empirical amino acid frequencies as well as four discrete gamma rates. Hundred bootstrap replicates were accomplished under the same model in RAxML with the rapid bootstrap option to estimate the support for internal nodes.

2.5 Cooperations and Contributions

The following people contributed and supported this thesis in different working methods due to their professional expertise:

Our technician Victoria Michael prepared the algal and parasitic cultures and isolated the genomic DNA and RNA of these cultures for the Illumina sequencers already before I started my PhD in the working group of PD Dr. Jörn Petersen (DSMZ). The cultivation of *Perkinsus* was learnt by V. Michael within an internship hosted by the working group of Christopher F. Dungan (Maryland Department of Natural Resources, Oxford, USA). The *P. olseni* strain PRA-181 was also kindly provided by C. F. Dungan. The cultivation and sequencing of the dinoflagellate *P. minimum* was accomplished in cooperation with the group of Prof. Dr. Irene Wagner-Döbler (Microbial Communication, HZI, Braunschweig). The Illumina sequencing was executed by the research group of Genome Analytics at the HZI in Braunschweig. The data management, assemblies and coverage analyses of the Illumina sequence datasets were performed by the IT specialist Michael Jarek (Genome Analytics, HZI, Braunschweig). The installation of a local blast platform, the training for NCBI uploads and the search for 16S rRNA with bioinformatic tools in used and unused reads of the Illumina datasets were

accomplished by our IT specialist Dr. Boyke Bunk (Microbial Ecology and Diversity Research, DSMZ). The immunocytochemical approach was instructed by Prof. Dr. Manfred Rohde (Head of the Technology Platform Electron Microscopy, HZI, Braunschweig). The preparation of the *Perkinsus* zoospores for the electron microscopy as well as the electron microscopic method was conducted by Prof. Dr. M. Rohde and his technical assistant Ina Schleicher. The phylogenomic analysis was accomplished in cooperation with Prof. Dr. Hervé Philippe (Département de biochimie et medicine moléculaire, l'université de Montréal, Canada). Phylogenetic reconstructions were computed by Dr. Henner Brinkmann (DSMZ). Sequence data for species of the gregarines, which were used for the species tree, were kindly provided by Prof. Dr. Guan Zhu (Department of Veterinary Pathobiology, Texas A&M University, Texas, USA).

3. Results

The newly established genomes and transcriptomes of the four key organisms *C. velia*, *V. brassicaformis*, *P. olseni* and *P. minimum* form the basis of all following analyses. Beforehand the quality of the assembled datasets was checked.

Three statistical values can be consulted to compare the Illumina datasets: the maximum contig/scaffold size, the N_{50} value and the average coverage value. In Figure 10 the quality parameters of the transcriptomic datasets of all four key organisms are plotted.

The highest contig/scaffold size is found in the chromerid *V. brassicaformis* (36,975bp) followed by the dinoflagellate *P. minimum* (26,514bp), the other chromerid *C. velia* (18,398bp) and finally the early-branching dinoflagellate *P. olseni* (7,822bp).

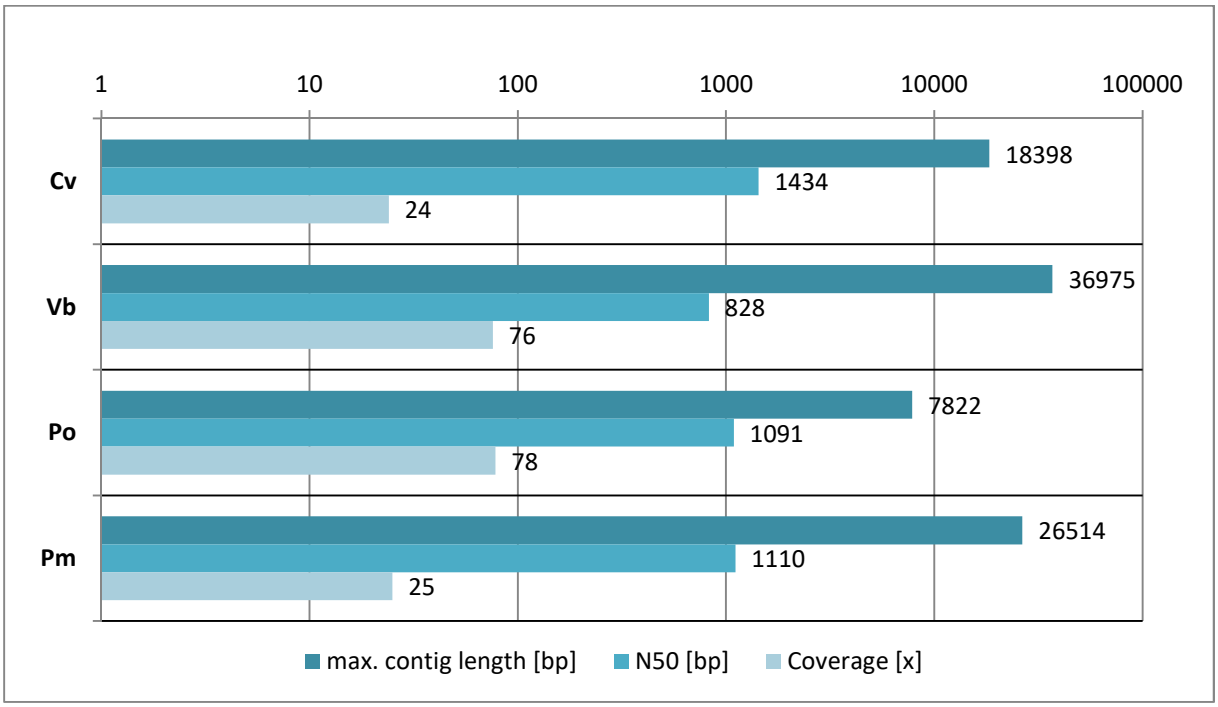


Figure 10: Overview and comparison of quality parameters for the four transcriptomic Illumina datasets of *C.velia* (Cv), *V. brassicaformis* (Vb), *P. olseni* (Po) and *P. minimum* (Pm). Shown are the maximum contig length, the N50 value and the average coverage value on a logarithmic scale.

The N_{50} value reflects that 50% of the *de novo* assembled transcriptome sequences are greater than or equal to the denoted contig/scaffold size. The higher the N_{50} value, the higher the rate of longer contigs/scaffolds and the better the assembly (Schmieder 2011). Referring to this definition, the 'best' dataset would be represented by *C. velia* (1,434bp).

The N_{50} value is often checked against the maximum contig/scaffold length of the dataset. The lowest range between the N_{50} value and the maximum contig/scaffold length is

represented by the oyster parasite *P. olseni* (N_{50} = 1,091bp; largest contig = 7,822 bp) and by the photosynthetic alga *C. velia* (N_{50} = 1,434bp; largest scaffold = 18,398bp), respectively.

The coverage value reflects the sequencing depth of a dataset. Thus, it displays the average frequency of how often contigs/scaffolds are covered within the draft genome. The higher the average coverage value, the better the assembled dataset. The best dataset is represented by *P. olseni* with a transcriptomic coverage of 78x (Figure 10) and a genomic coverage of 140x (chapter 0.43.1.4, Table 3).

In the case of the non-axenic chromerid cell cultures, the values for contig sizes and coverage values are biased by bacterial sequences. Even the transcriptomes show indications of bacterial contaminants. This observation is well documented by the GC plot of the *V. brassicaformis* transcriptome, where the GC content do not follow a conventional normal distribution, but displays a bimodal graph that is often observed for metagenomic datasets (Figure 11A). In contrast, the demonstrably axenic cell culture of the parasitic dinoflagellate *P. olseni* shows a clearly normal distribution in its GC plot (Figure 11B).

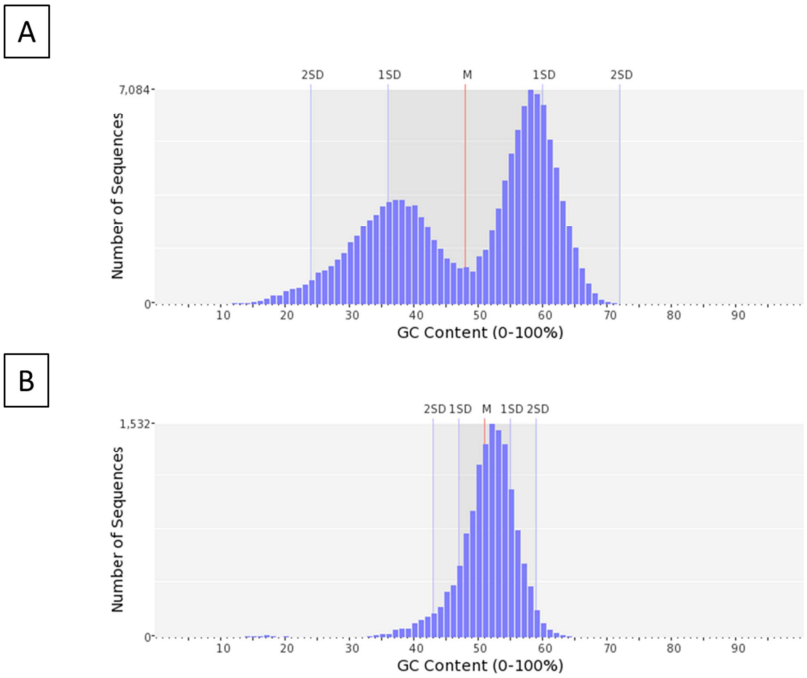


Figure 11: GC plots of transcriptomic datasets. A) Bimodal distribution of the GC content in the transcriptomic dataset of the non-axenic chromerid *V. brassicaformis* versus B) normal distribution of the GC content in the transcriptomic dataset of the axenic *P. olseni* strain. Analysis and visualization by Prinseq (Schmieder, Edwards 2011).

3.1 Plastid Evolution in Alveolata

3.1.1 Species Tree and Its Relevance for Plastid Evolution

The superensemble of alveolates is comprised of the Ciliata, Dinoflagellata and Apicomplexa (Adl et al. 2012). The distribution of the plastid organelle in these phyla, especially in the phylum Apicomplexa, is heterogenous due to a parasitic non-photosynthetic life style and the absence of this organelle in some species or whole groups (chapter 1.4.3/1.4.5; Figure 12).

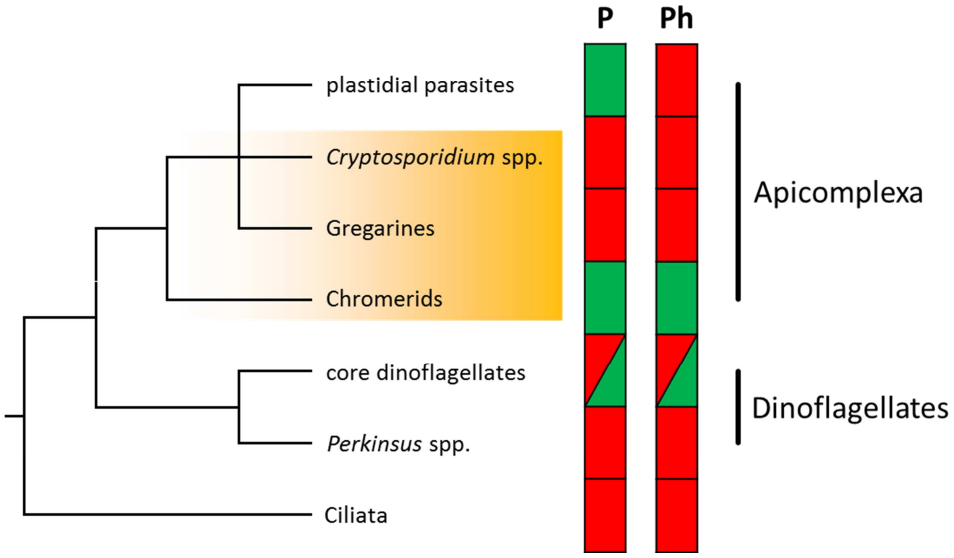


Figure 12: Presence/absence of plastids (P) and photosynthesis (Ph) in alveolates. Presence is marked by green, absence by red. The basal apicomplexan lineages are highlighted by a yellow box.

In nuclear phylogenies representing the eukaryotic host cell, the precise positioning of early branching apicomplexan lineages – the photoautotrophic algae *C. velia* and *V. brassicaformis*, as well as the aplastidial genus *Cryptosporidium* and some aplastidial gregarines – among the apicoplast bearing parasites is not resolved (chapter 1.4.3/1.4.5). However, in a recent phylogenomic approach including the chromerids and also the genus *Cryptosporidium*, there is an indication that the photosynthetic algae are the earliest branching apicomplexan group (Janouskovec et al. 2015, Figure 12). This result would either support a common photosynthetic origin in all apicomplexan lineages or two independent gains of plastids in chromerids and plastid-bearing apicomplexan parasites. Thus, a reliable reconstructed alveolate species tree also including the gregarines would give information about the exact position of the early branching sisters of the plastidial apicomplexan parasites and would additionally serve as a backbone for further plastid protein phylogenies in order to address the question of plastid gain and plastid loss among the Apicomplexa.

Therefore, a phylogenomic study was performed that included our newly established transcriptomes for *C. velia*, *V. brassicaformis*, *P. olseni* and *P. minimum* as well as two *Cryptosporidium* species and four gregarines (kindly provided by Dr. Guan Zhu, Texas A&M University). The reconstructed alveolate species tree is based on a concatenated dataset of more than 150 orthologous nuclear genes (Figure 13, Figure S3) and thus serves as a reliable backbone for further evolutionary analyses of their organelles.

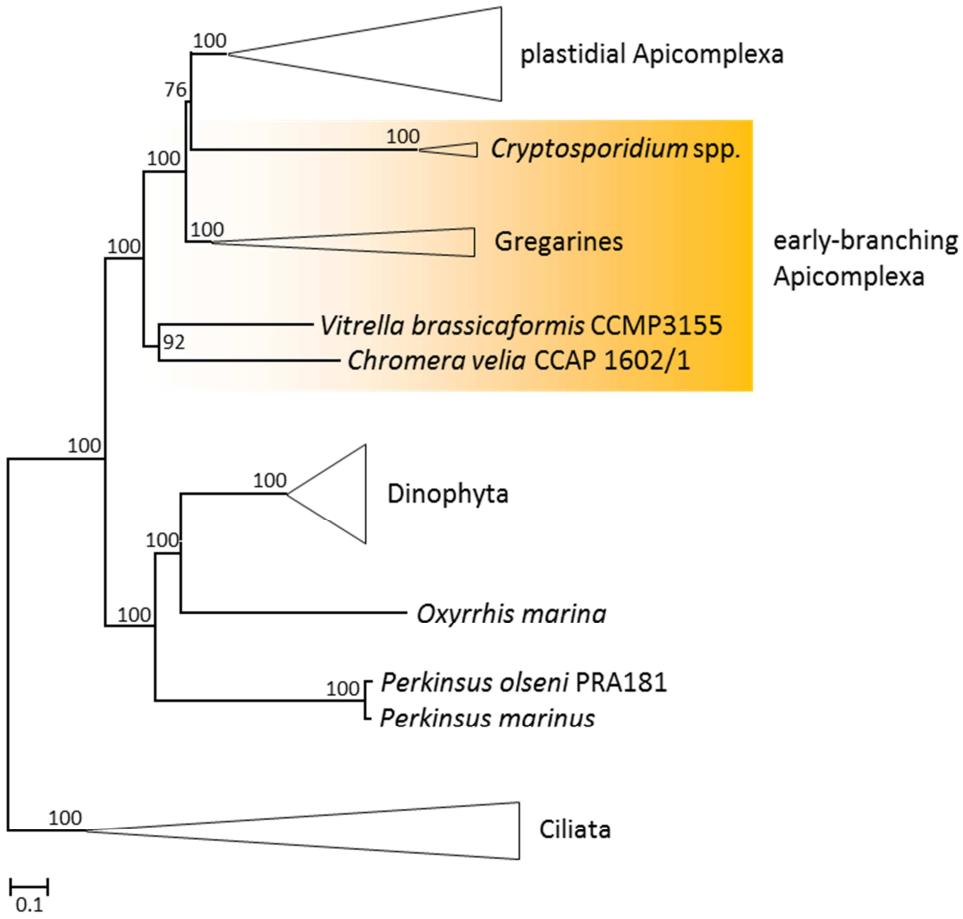


Figure 13: Phylogenomic tree of 46 alveolate species with 30,065 aa positions based on a CAT-GTR tree reconstruction model. Ciliates form the outgroup. Clearly defined groups are collapsed to triangles. The bootstrap values are according to the jackknife resampling method. The basal apicomplexan species are highlighted in a yellow box. The complete phylogenomic tree is shown in Figure S3.

The phylogenomic analysis revealed well defined alveolate subgroups with 100% bootstrap support (BS, Figure 13, Figure S3). The deepest branch of alveolates is represented by the ciliates supporting previous phylogenies (Figure 13; Gajadhar et al. 1991; Cavalier-Smith 1993; Fast et al. 2002). The heterotrichous species *Condylostoma magnum* forms the earliest branch of this phylum corresponding to the position of the Heterotrichea in the ciliate tree (Gao et al. 2016). The subphylum Intramacronucleata is divided into two main groups

referred to the subdivision of Gao et al. (2016): the 'CONthreeP' group (*Ichthyophthirius*, *Tetrahymena*, *Paramecium*, *Chilodonella*) and the 'SAL' group (*Stylonychia*, *Oxytricha*, *Favella*). Both ciliate subtrees are monophyletic with 100% BS (Figure S3).

The dinoflagellate subtree is monophyletic with a serial branching order of the parasitic *Perkinsus* spp. and *Oxyrrhis marina* at the earliest positions (Figure 13). The core dinoflagellates are monophyletic spanning internal monophyletic groups of the Kareniaceae (*Karenia*, *Karlodinium*; 100% BS), the Suessialeans (*Symbiodinium*; 100% BS) and the Gonyaulacoids (*Alexandrium*, *Pyrodinium*, *Lingulodinium*; 100% BS) (Figure S3).

Within the monophyletic group of Apicomplexa (100% BS) the photosynthetic algae *C. velia* and *V. brassicaformis* represent the earliest branching lineages forming a monophyletic group with 92% BS, followed by a serial branching order of the monophyletic groups of gregarines and *Cryptosporidium* spp., where the position of *Cryptosporidium* spp. is supported more weakly with a 76% BS (Figure 13, Figure S3). The plastidial apicomplexan parasites respresented by the Conoidasida and Aconoidasida, as well as the subclasses Piroplasmida and Haemosporidia of the Aconoidasida, are monophyletic (100% BS) as expected (Adl et al. 2012, Figure S3).

The chromerids as the deepest branch indicate that the Apicomplexa have recruited their plastid in a common ancestor of this phylum and as *argumentum e contrario* that the aplastidial lineages have lost their plastid secondarily if the species tree is congruent with the evolution of the apicoplast. The puzzle of the early evolution of the Apicomplexa is solidly resolved with this study. The deep branching of the chromerids indicates that the origin of the Apicomplexa is an ancient event. Since the monophyly of the chromerids is not supported with 100% an alternate position of the two species in serial order at the basis of all other apicomplexan lineages is possible. Likewise, the monophyly of the non-plastidic gregarines and *Cryptosporidium* spp. cannot be rejected with respect to the bootstrap value of only 76% of the *Cryptosporidium* branch (Figure 13).

3.1.2 The Monophyly of the Apicoplast and the First Evidence of Plastid Loss

In order to test the hypothesis of plastid loss in both apicomplexan aplastidial lineages (*Cryptosporidium* spp. and the gregarines) it was tested if phylogenies of apicoplast genes would reveal a monophyletic relationship between all apicomplexan species as demonstrated by the species tree (Chapter 3.1.1, Figure 13). Therefore, plastid-specific markers were selected, which are not involved in the photosynthesis apparatus to maximize the apicomplexan taxon sampling including also the parasitic species (e.g. *Plasmodium* and *Toxoplasma*). The seven genes of the plastid-sustaining MEP pathway for isoprenoid biosynthesis were selected, because this pathway was defined as the Achilles' heel of apicomplexan parasites (Jomaa et al. 1999; Yeh, DeRisi 2011; Chapter 1.5.2).

Frequently, the very fast evolving apicomplexan species bias the phylogenetic signal generating phylogenetic noise that leads to tree reconstruction artifacts such as the long-branch attraction artifact (LBA). Accordingly, the preliminary phylogenetic reconstructions of the MEP proteins revealed only three out of seven proteins as suitable markers for further analyses (DXS, HDS and HDR) with enough amino acid alignment positions. In Figure 14 the three tree reconstructions are displayed.

The **DXS** tree (Figure S5) shows that the protein sequences are of alpha-proteobacterial origin. Furthermore, the CASH lineages (the dinoflagellates excluded with 86% BS) are supported to contain a "red" complex plastid by forming a monophyletic group with the rhodophytes (86% BS). *G. theta* (cryptophyte) and *E. huxleyi* (haptophyte) are branching weakly supported close to the stramenopiles. The monophyly of the apicoplast is well supported (70% BS) and in a separate analysis of the red subtree the support of the apicoplast even reached 90% BS (Figure 14; Figure S4). The photoautotrophic algae, *C. velia* and *V. brassicaformis*, are located at the basis of the apicoplast group forming a monophyletic chromerid group with 74% BS (Figure 14).

The **HDS** phylogeny of Viridiplantae (Chlorophyta, Streptophyta) and complex algae (Figure 14, Figure S6) shows a chlamydial origin of this gene. In contrast, the red algal and glaucophycean HDS sequences of this plastid marker branch in another part of the tree and are of cyanobacterial origin (data not shown). Thus, the original HDS gene must have been replaced by the chlamydial equivalent in a common ancestor of the green and the CASH lineages indicating a green algal ancestry of the HDS in the complex algae with red plastids. The Myzozoa branch serially as the sister groups to all other eukaryotic sequences and thus does not support a common origin of their plastids (Figure 14). The cryptophyte *G. theta* is associated to a group of chlorophytes with low bootstrap support, whereas the two sequences

of the haptophyte *E. huxleyi* are nested within the stramenopiles (100% BS, Figure 14, Figure S6). The Apicomplexa are monophyletic with 90% BS. Again, both chromerids form a monophyletic group at the basis of apicomplexan parasites (88% BS, Figure 14).

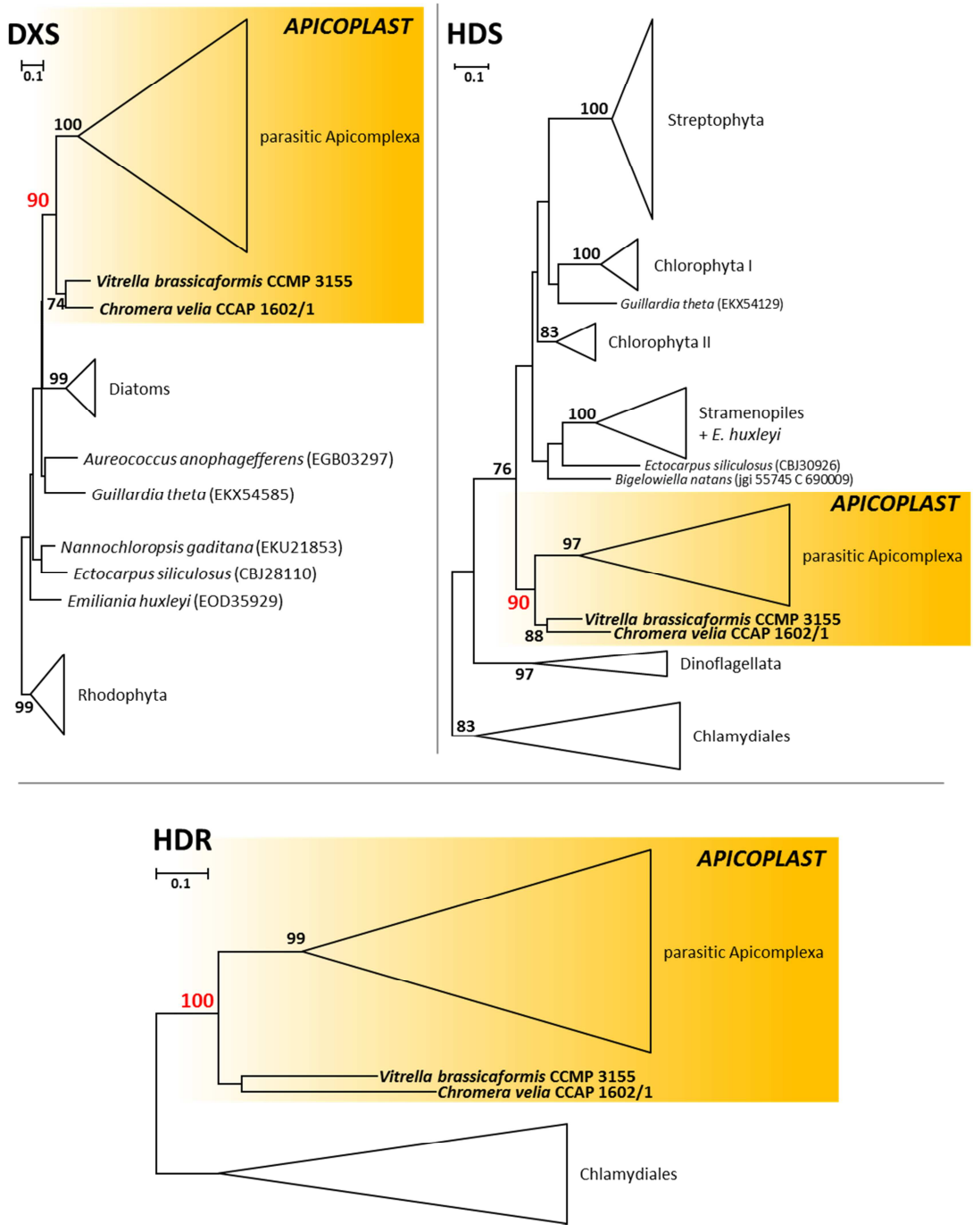


Figure 14: Phylogenetic maximum likelihood RAxML trees (LG+F+Γ4) for the three MEP proteins DXS, HDS and HDR. Well defined taxonomic groups are collapsed to triangles. Sequences representing the apicoplast are highlighted by a yellow box. Only Bootstrap values ≥ 70% are shown. The bootstrap values for the monophyly of the apicoplast are highlighted in red. The original trees are documented in the supplements (Figure S4; Figure S6 and Figure S7).

The original cyanobacterial **HDR** gene was also replaced by a chlamydial equivalent, but only in the apicomplexan species (Petersen et al. 2014). The chlamydial subtree is shown in Figure 14. The photosynthetic chromerids form a monophyletic group with the parasitic apicomplexan lineages (100% BS) supporting a common origin of their plastids. The photosynthetic species, *C. velia* and *V. brassicaformis*, are monophyletic even if the statistical support is low (36%BS, Figure S7), but they share a common feature in terms of the gene structure (Figure 15).

The HDR gene is merged at its 5´- end with a gene of the Calvin cycle, the sedoheptulose-1,7-bisphosphatase (SBP, Figure 15; Petersen et al. 2014), to build a fusion protein (*C. velia*: 730 aa, *V. brassicaformis*: 759 aa). Both genes are interrupted by numerous introns. The 15 introns of *C. velia* are larger than the 21 very short introns of *V. brassicaformis*. The proteins dispose of a ~40 aa hinge-region, which connects the two proteins with each other (Figure 15).

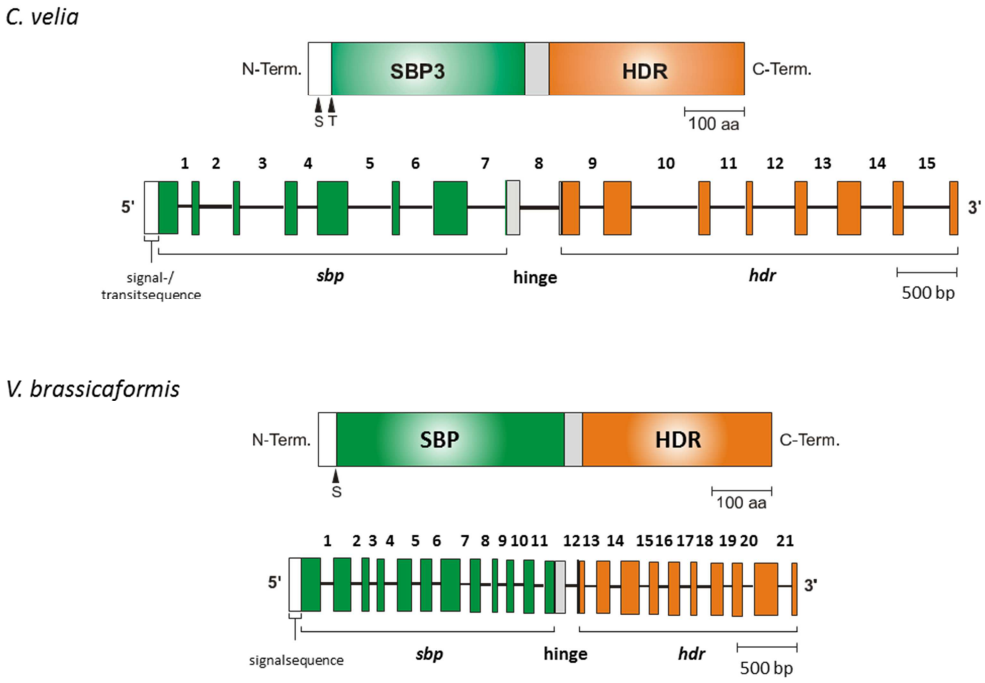


Figure 15: Shared fusion protein in the photosynthetic apicomplexan species, *C. velia* and *V. brassicaformis*. The N-terminal SBP of the calvin cycle is shown in green, the C-terminal HDR of the MEP pathway is marked in orange. Presequences are white; the intermediate hinge region is grey. Cleaving sites of the signal- and transit-sequences are indicated by arrowheads. The graphic incorporates the mature protein sequences (top graphic) as well as the genomic sequence interrupted by serial numbered introns. Abbreviations are in compliance with the text.

Additional suitable markers to study plastid evolution are those of the protein import machinery in complex red plastids. The outermost membranes of the complex plastids regulate the transport by the highly conserved symbiont-specific ERAD-like machinery (SELMA; Hempel et al. 2009). This import cascade was described to be “recycled” from the

endoplasmatic-reticulum-associated protein degradation (ERAD) system. For example, the plastidial SELMA *cdc48* gene is a paralog to the cytosolic ERAD-form (Bolte et al. 2011).

The phylogeny of the **CDC48** protein is shown for the red subtree of the CASH plastids (Figure 16). A monophyletic group of rhodophytes and cryptophytes forms the basis of this tree. The only two dinoflagellate CDC48 sequences in this study (*K. brevis* and *K. micrum*) cluster together with the sequence of the haptophyte *E. huxleyi* (100% BS). This result is not surprising as both species contain plastids they once acquired by a tertiary endosymbiosis with a haptophyte alga (Chapter 1.4.5). A big cluster of stramenopile sequences are the sister group to the monophyletic group of Apicomplexa (98%, Figure 16). The chromerid sequences are also monophyletic (94% BS) and branch together with the apicomplexan suborder Eimeriorina (47% BS). This group is the sister group to the apicomplexan Aconoidasida (93% BS).

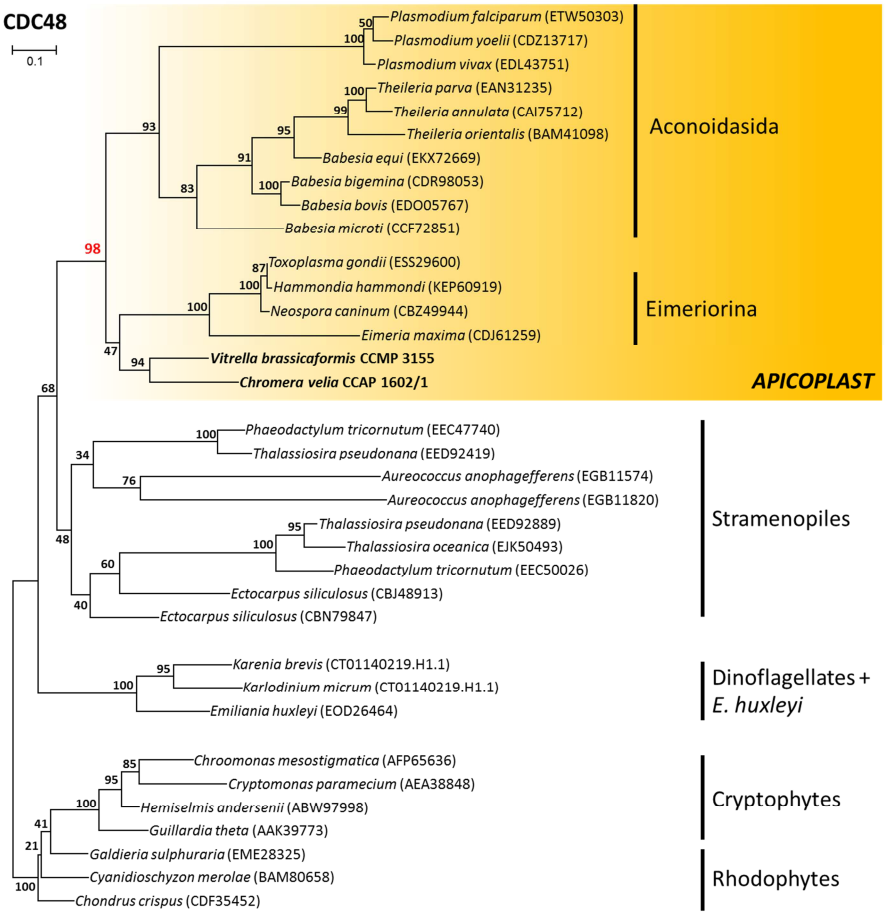


Figure 16: Phylogenetic maximum likelihood RAxML tree (LG+F+Γ4) based on 35 eukaryotic species incorporating 553 positions for the SELMA protein CDC48. Sequences representing the apicoplast are highlighted by a yellow box. Sequences established within this thesis are highlighted in bold. The bootstrap value for the monophyly of the apicoplast is marked in red.

The transport of proteins through the two innermost membranes of the plastid organelle is regulated by the highly conserved TIC/TOC import system (Schnell et al. 1997). Unfortunately, for phylogenetic reconstructions the number of amino acid positions of these proteins is often limited due to highly divergent alignments and their short sequence length (on average 300 – 350 aa), thus the trees are prone to artifacts. However, the **TIC20** protein was chosen as a suitable marker because it was previously shown to be essential for the growth of the apicomplexan parasite *T. gondii* (van Dooren et al. 2008). The tree was already calculated and shown in the supplements of our publication about the Rhodoplex hypothesis (Petersen et al. 2014; Figure S8). Corresponding to its plastidial context, the TIC20 protein is of cyanobacterial origin. The position of distinct taxonomic groups is only weakly supported. The dinoflagellate *P. minimum* contains five copies of this plastidial protein and all copies cluster together with stramenopile sequences. In *P. olseni* this protein seems to be absent. The other CASH lineages are distributed all over the tree with no clear evolutionary resolution, whereas the Apicomplexa builds a monophyletic group with weak support (41% BS). As expected *C. velia* is the deepest branch of all other apicomplexan species. However, *V. brassicaformis* does not cluster with the other apicomplexan species, but among the other CASH members, far from its closest relative *C. velia*. Likewise noticeable is the long branch of *C. velia* in comparison to the other apicomplexan species. Perhaps, some lineages experienced a kind of evolutionary acceleration in their primary plastid protein import system (species with long branches), which results in these basically insignificant support values and disordered placements of CASH species. Thus, this phylogenetic marker is not suitable for general conclusions about plastid evolution.

In summary, all plastid phylogenies presented in this study indicate a common origin of the apicoplast (partial support by the TIC20 phylogeny). In all analyses (Figure 14/Figure 16/Figure S8), the apicomplexan species form a monophyletic group, which is in congruence with the species tree. Thus, these results clearly demonstrate that the plastid of the chromerids can be referred to as a photosynthetic active apicoplast. Additionally, due to the monophyly of the apicoplast with the photosynthetic chromerids as the deepest branch, the analyses provide the first clear evidence for plastid loss in the apicomplexan genus *Cryptosporidium* and the gregarines when compared to the established species tree (Chapter 0.43.1.1, Figure 13).

3.1.3 Transcriptomics of the Early Branching Dinoflagellate *P. olseni*

The KAAS (KEGG automated annotation server; Moriya et al. 2007) platform is a useful tool for a rapid identification and assignment of sequence information in a draft dataset. In this case I wanted to get a general idea of how much information can be extracted from the newly established transcriptomes. In Figure 17 the general information content is illustrated in a compressed form. Again, the transcriptomic dataset of *P. olseni* emerges as the best dataset because KEGG orthologs (KOs) were assigned to a high number of sequences compared to the total number of sequences (Figure 17, 4711/14001 sequences). Likewise, a large number of sequences in the *C. velia* transcriptome could be annotated to appropriate KOs (Figure 17, 8642/46868 sequences). Because the transcriptomic dataset of *V. brassicaformis* is biased due to bacterial contaminations, it was excluded from further analyses. In general, the results of the KO annotations are classified into three categories: *genetic information processing*, *metabolism* as well as *signaling and cellular processes* (Figure 17). The output reflects well the discrepancy between the autotrophic species and the parasitic organism.

Obviously, less orthologous genes could be identified within the dataset of the parasitic dinoflagellate *P. olseni*, indicating the partial dependency on the host (Figure 17). In contrast, the photoautotrophic dinoflagellate *P. minimum* seems to have an expanded contingent of positive annotations, likely due to the known gene and sequence accumulation in autotrophic dinoflagellate genomes (e.g. Wisecaver et al. 2013). The metabolic category is the most represented category in all three organisms (Figure 17). Interestingly, seven sequences of the *P. olseni* transcriptome were annotated as photosynthetic proteins (Table S4). Due to the assumption that *Perkinsus* spp. does not harbor a plastid, these annotations were tested via manual blastx searches whereby the results could not be attested. Most of these sequences were assigned to hypothetical proteins, which were first described in a close relative of *P. olseni*, *P. marinus* (Table S4). Thus, the annotation accuracy of the KAAS platform is limited. The program relies on sequence similarity and BLAST bit scores alone and thus depends on the bit score cutoff. In order to maximize the hit rate of the highly diverse alveolate sequences I changed the bit score cutoff from 60 to 30. Hence, all BLAST hits with a bit score lower than 30 were removed. Unfortunately, lowering the bit score cutoff accumulates false positives like these detected photosynthetic or even bacterial proteins in the axenic *P. olseni* strain (data not shown). Otherwise it increases the chance to identify highly diverse sequences. Furthermore, to guarantee the calculating speed of KAAS, the sequences are only

compared to a selected set of genomes (KEGG GENES database), which could represent another error source for the absence of the assignment of sequences.

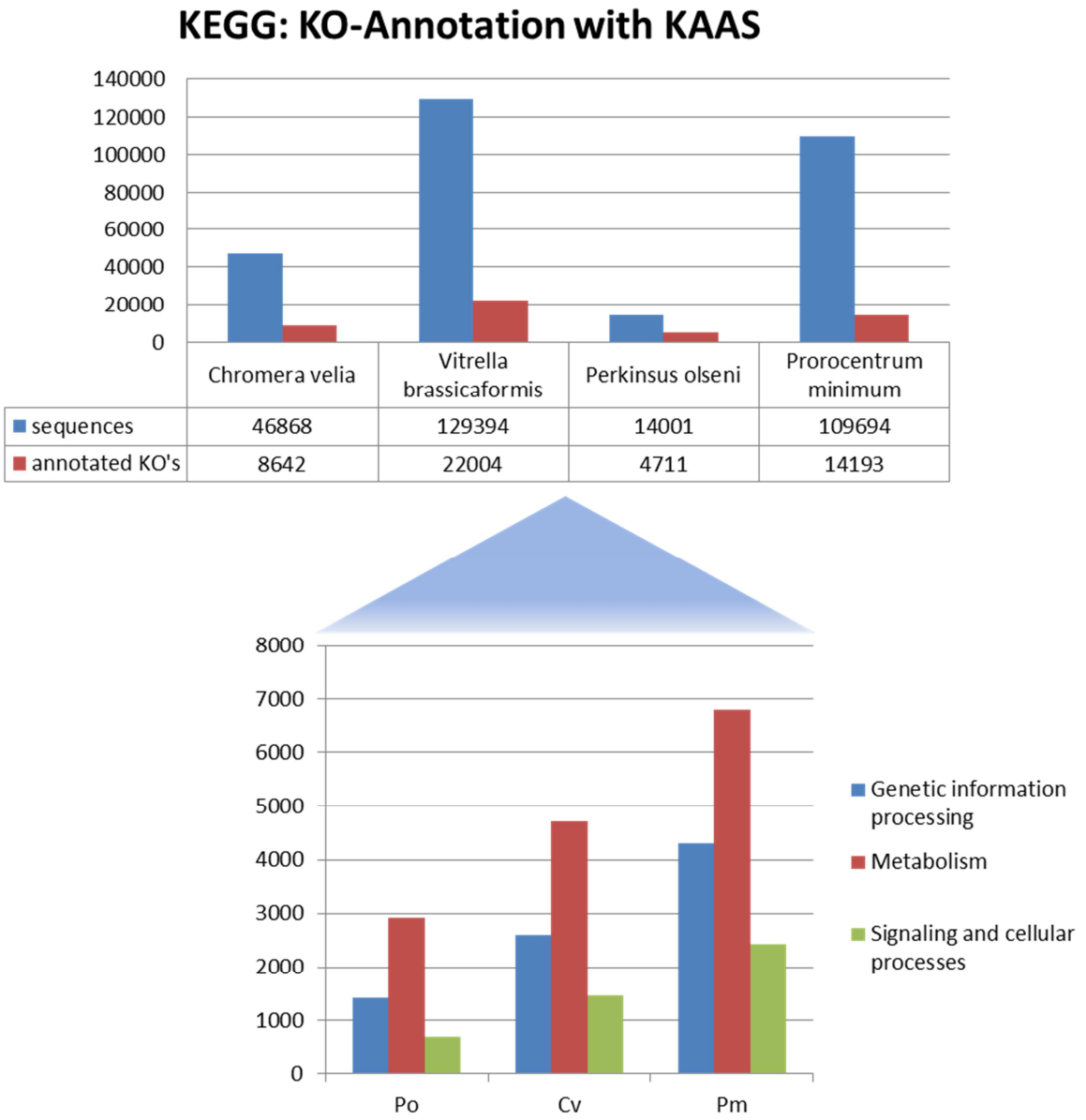


Figure 17: KO-Annotation of the algal transcriptomes with KAAS (Moriya et al. 2007). The first comparative statistic reflects the number of sequences classified to the KEGG Orthology (KO) database in relation to the total number of sequences in the respective transcriptome. The second statistic shows the number of annotated sequences within the classification groups: "Genetic information processing", "Metabolism" and "Signaling and cellular processes". Abbreviations: Po, *Perkinsus olseni*; Cv, *Chromera velia*; Pm, *Prorocentrum minimum.*

The pipeline of the annotation program Blast2GO (Conesa et al. 2005) is more accurate, because it generates a meaningful blast table containing all blast results for each sequence of the respective transcriptome. The program is based on blastx searches against the current non-redundant *nr*-database of NCBI, regardless of a bit score cutoff. In addition, an implementation of the InterPro program allows deducing protein domains and motifs from the nucleotide sequences. Both information sources are combined to be compared with the Gene

Ontology (GO) database in order to retrieve the GO terms. In the annotation step, the obtained GO terms are assigned to the respective query sequences based on an annotation score. Frequently, the EC numbers of the annotated enzymes are denoted. A basic Blast2GO analysis was accomplished with the smallest dataset of the studied organisms, the transcriptome of *P. olseni*. The transcriptomes of *V. brassicaformis* and *P. minimum* were not calculated because of the bacterial contaminations in *V. brassicaformis* and the large amount of sequences in both transcriptomes would lead to long computation times. The transcriptome of *C. velia* is problematic, because the first 99 sequences exceed the maximum scaffold length of 8000 bp to be sent to the NCBI Blast server. These missing data would distort the total outcome of the analysis.

The transcriptome of *P. olseni* contains in total 14001 sequences for which the program displayed 10663 sequences with blast hits. The remaining 3338 sequences showed no significant similarity to any sequence of the *nr*-database of NCBI (Figure 18). The shorter the sequences the more frequent no blastx result was available. After all four analytical steps (blast, interpro, mapping and annotation) the data distribution is composed of (1) 3943 sequences that have a significant blast hit, but could neither be mapped nor annotated to the GO database, (2) 1153 sequences with a positive mapping to the GO database, but no annotation to a specific GO term and finally (3) 5567 sequences which could be actually annotated (Figure 18). Additionally, enzyme codes could be added to 1476 sequences.

In Figure 18 the top hits of species assigned via the blastx search are recorded. 10611 sequences of the 10663 positive blast results are related to eukaryotic sequences, whereas 52 sequences display a bacterial origin, mostly represented by γ-proteobacteria. Since *P. olseni* is an axenic cell culture, these species annotations could reflect gene replacements or gene acquisitions by horizontal gene transfer from bacteria. Otherwise, if the underlying similarity is very low, it could be a false positive.

Most of the eukaryotic sequences (10369 of 10611, Figure 18) were alveolate blast hits, with a high number of *P. marinus*, the closest relative of *P. olseni*. The transcriptome of *Perkinsus marinus* is well-defined and studied by EST projects (e.g. Joseph et al. 2010), thus the abundance of *P.marinus* sequences is not astonishing. Both *Perkinsus* species must be very close related and similar to each other concerning their genetic information content to obtain such a high number of blast top hits for *P. marinus*. The fraction of annotated dinoflagellate sequences within the sequences of eukaryotic origin is 96.2%, 1.4% were assigned to apicomplexan species, thus in summary 97.6% are related to myzozoan species. Accordingly, 3532 sequences classified in the NCBI database as hypothetical proteins of *P. marinus*

("PMAR") were detected in the transcriptomic dataset of *P.olseni*. 94 sequences were assigned to other hypothetical proteins e.g. of *Guillardia theta* ("GUITH"), *Emiliania huxleyi* ("EMIHU") or *Toxoplasma gondii* ("TG").

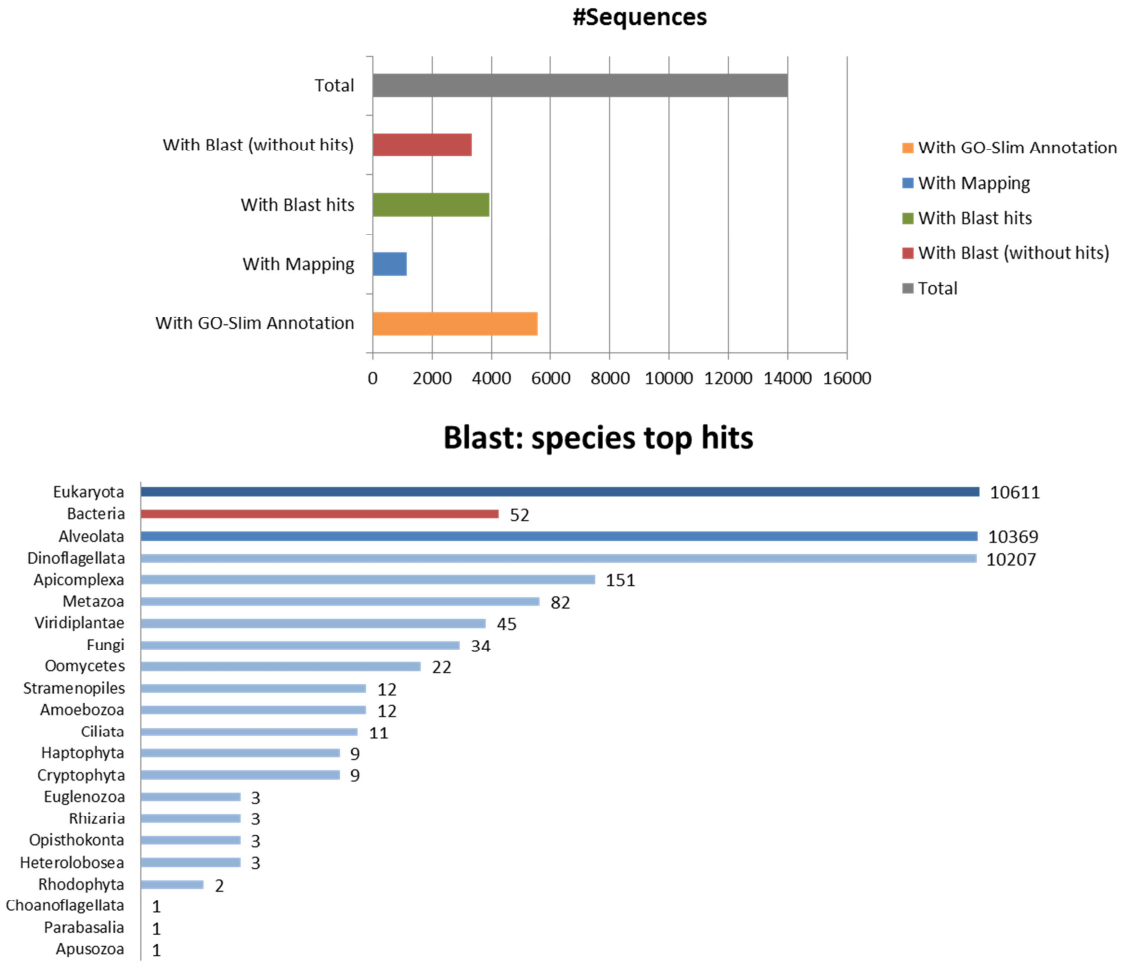


Figure 18: Two different output statistics of the Blast2GO analysis (Conesa et al. 2005) for the *P. olseni* transcriptome. Illustrated is the data distribution after each Blast2GO analysis step as well as the species distribution after the blastx analysis showing only the top hits. Represented in detail are the blast results with eukaryotic hits, whereas bacterial hits are marked and summarized by the red color.

Like in the KAAS platform, the annotated sequences of the transcriptome are classified into three categories in Blast2GO: "cellular component" (CC), "biological process" (BP) and "molecular function" (MF). The output for these categories can be accessed for different levels of the GO term hierarchy. Level 8 was chosen for the CC category showing the percentages of the respective annotated enzymes in different cellular sites (Figure 19).

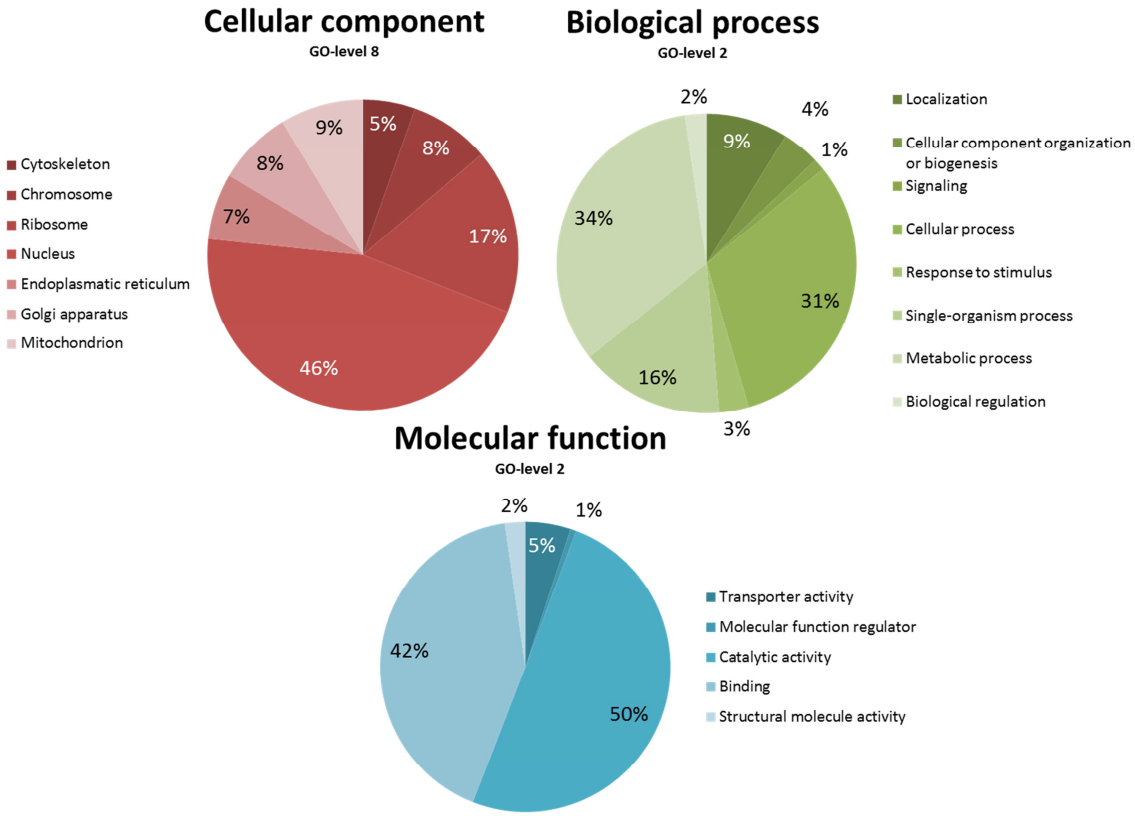


Figure 19: GO-annotations of the dinoflagellate transcriptome of *P. olseni* performed by Blast2GO (Conesa et al. 2005). Shown are different GO-levels: for the category "Cellular component" level 8 and for the categories "Biological processes" and "Molecular function" level 2, respectively. The number of annotated sequences within this category are labeled by percentage of the corresponding subcategory.

Most of the proteins are located in the nucleus or are assigned to ribosomes (46% and 17%, Figure 19). A fraction of 9% is related to proteins of the mitochondrion, whereas no annotations were associated with plastid structures. For the two other categories the more general GO level 2 was chosen as a quick overview of the processes and functions these proteins are involved. Not surprisingly, in the category "Biological Processes" almost two-thirds of the proteins are involved in cellular and metabolic processes (31% and 34%). Both processes are roughly equal represented, whereas in the KO annotation the metabolism category was most abundant (Figure 17). In most instances the annotated proteins resume a catalytic (Figure 19, 50% with big amounts of hydrolases and kinases, GO level 5) or binding function (42% with RNA binding > DNA binding, GO level 5). The high level of RNA binding proteins indicates that in *P. olseni* RNA processing and post-transcriptional gene regulation plays a central role. Kinases are important to transmit signals within the cell, thus the parasitic organism can react rapidly to different stimuli or to changes in its environment.

3.1.4 The "Perkinsuplast": A Plastid Without a Genome in the Dinoflagellate Genus *Perkinsus*

The enigma of the presence of a plastid in the basal dinoflagellate genus *Perkinsus* is still not unraveled. Multiple publications addressed to this question (e.g.: Grauvogel et al. 2007; Stelter et al. 2007; Teles-Grilo et al. 2007; Matsuzaki et al. 2008; Joseph et al. 2010; Fernandez Robledo et al. 2011), but could not clearly demonstrate the presence or absence of a plastid. In this part of the thesis, it was tried to detect a hypothetical plastid genome via diverse analytical techniques *in silico*. Furthermore, a possible plastid structure was searched via microscopic applications *in vitro*.

16S rRNA-based Analysis

Primarily, local blastn searches in the *P. olseni* metagenome with appropriate plastid-specific 16S rRNA query sequences as a marker for a possible chloroplast genome were performed. Close relatives of *Perkinsus* were chosen as eukaryotic query sequences, e.g. the apicomplexan species *Theileria annulata*, the dinoflagellate species *Lingulodinium polyedrum*, the chromerid species *C. velia* or the stramenopile *Phaeodactylum tricornutum*. The 16S rRNA of plastids is highly conserved, thus a highly similar 16S rRNA sequence of *P. olseni* in the genomic dataset has to be found. However, the best hits were similar to eukaryotic 18S rRNA, but not to 16S rRNA. Furthermore, a 'naive Bayesian classification' of the assembled Illumina genome was conducted with a 16S rRNA training set of the RDP classifier (Wang et al. 2007). 20 sequences were assigned to the group of Cyanobacteria/Chloroplast, but were not verified by single blastn searches in NCBI. Additionally, we were looking for 16S rRNA sequences in the unused reads (51,905,313 sequences) of the Illumina genomic dataset assembly. The software SortMeRNA (Kopylova et al. 2012) checked these unused reads against a database containing bacterial 16S rRNA sequences. 405,375 reads (0,781%) were detected as possible 16S rRNA sequences. This result was again tested with SortMeRNA against a database containing eukaryotic 18S rRNA, which is highly similar to 16S rRNA. 86% of these sequences (350,062/405,375 reads) were actually classified as eukaryotic 18S rRNA. The remaining 55,313 reads, which were previously identified with the bacterial 16S rRNA database, were classified with the RDP classifier. Only 7,474 out of the 55,313 reads were assigned to bacterial and archaeal categories. Thirteen hypothetical chloroplast 16S rRNA sequences with an average length of only 58bp could be detected. These sequences were checked with blastn searches within the BLAST-Server of NCBI, but did not score any 16S rRNA sequences. Accordingly, the results

of the classifier, which is usually used for the assessment of 16S-rDNA amplicon libraries, had to be treated with care, especially due to the very short lengths of the reads. In conclusion, neither the assembled Illumina genome (140-fold coverage) nor the unused reads revealed any hints for the presence of the 16S rRNA gene.

Coverage-based Analysis

The relative abundance of accessory organelles in the cell is much higher than the frequency of the nucleus. Since the nucleus is present only once per cell, organelles like mitochondria and plastids can be found multiple times. Thus, their genome sequences are sequenced more often than nuclear sequences. Therefore, plastidial and mitochondrial sequences are detectable by their high coverage value in the Illumina genomic datasets.

The high quality genomic dataset of *C. velia* was used as a reference. Concerning the (1) nuclear location, nuclear-encoded MEP plastid genes were chosen as test sequences; for the (2) mitochondrial location cytochrome oxidase 1 and cytochrome b sequences were used; and for the (3) plastid location, the already annotated *C. velia* plastome (Janouskovec et al. 2010) was utilized as a blast query. In the *C. velia* dataset with a 17-fold genome coverage, mitochondrial sequences as well as plastome contigs could be detected easily via local BLAST analyses with a 150-fold coverage of the mitochondrial genome and a ~ 100-fold coverage of the plastome (Table 3). Likewise, in the genomic dataset of *P. olseni* (140-fold coverage), the mitochondrial sequences were recorded with coverage values of around 15,000x (Table 3).

Table 3: Genome coverage values of *P. olseni* and the chosen reference organism *C. velia* for nuclear, mitochondrial and plastid test sequences. The abbreviation 'NA' ('not available') marks missing data.

Genome	*C. velia* coverage	*P. olseni* coverage
nuclear	~17x	~140x
mitochondrial	~150x	~15.000
plastid	~100x	NA

Referring to my working hypothesis that if there exists a putative chloroplast in *P. olseni* that contains DNA, one would assume a minimal coverage value approximately as high as the average nuclear genome coverage if it occurs at least once per cell. As aforementioned we found no evidence for the presence of plastid 16S rRNA in the genomic assembly as well as in the unused reads. Local BLAST analyses with other typical plastome sequences of essential plastid pathways (Chapter 1.5.2; Table S5) resulted in no hits or in hits comparably to nuclear coverage values and the presence of introns, thus indicating a nuclear localization of these genes.

Comparison of Plastome-encoded Genes in Genomes of Protistan Species

In the next approach, genes were identified, which could be still encoded on a hypothetical plastid genome in *Perkinsus* sp. and, which are thus essential for the maintenance of this genome. To proof our working hypothesis of a completely reduced plastid in this genus, *Perkinsus* was compared with another close parasitic relative, the malaria parasite *P. falciparum* that is known to harbor a reduced plastid: the apicoplast. According to Sanchez Puerta et al. (2005) the localization of the apicoplast genes was compared with those of primary (*Cyanophora paradoxa*, *Chondrus crispus*, *Arabidopsis thaliana*) and secondary plastomes (*Guillardia theta*, *Emiliania huxleyi*, *Thalassiosira pseudonana*, minicircles of Dinophyta) and visualized in a Venn diagram (Figure 20; Accessions see chapter 2.2.3).

The Venn diagram (Figure 20) shows that *P. falciparum* shares 23 of its apicoplast genes with other protistan plastid genomes. 21 of the apicoplast encoded genes are involved in the plastid transcription and translation apparatus, the remaining three are assigned to cell processes (*clpC*), transport (*sufB*) or a hypothetical function (*orf91*).

The gene expression apparatus should only be maintained in the plastid if there are remaining proteins encoded on the plastome. Thus, genes which are indispensable for essential functions of the heterotrophic plastid were included into a new Venn diagram and genes for the transcription and translation categories were omitted. In addition, genes which were assigned to the 'photosynthesis' category of Sanchez-Puerta et al. (2005) were disregarded in this non-photosynthetic parasitic context. Furthermore, genes with a hypothetical function which are often synapomorphic for distinct species were skipped. In Figure 21 the new Venn diagram of this reduced dataset is shown, simply compared with reference species harboring a primary plastid.

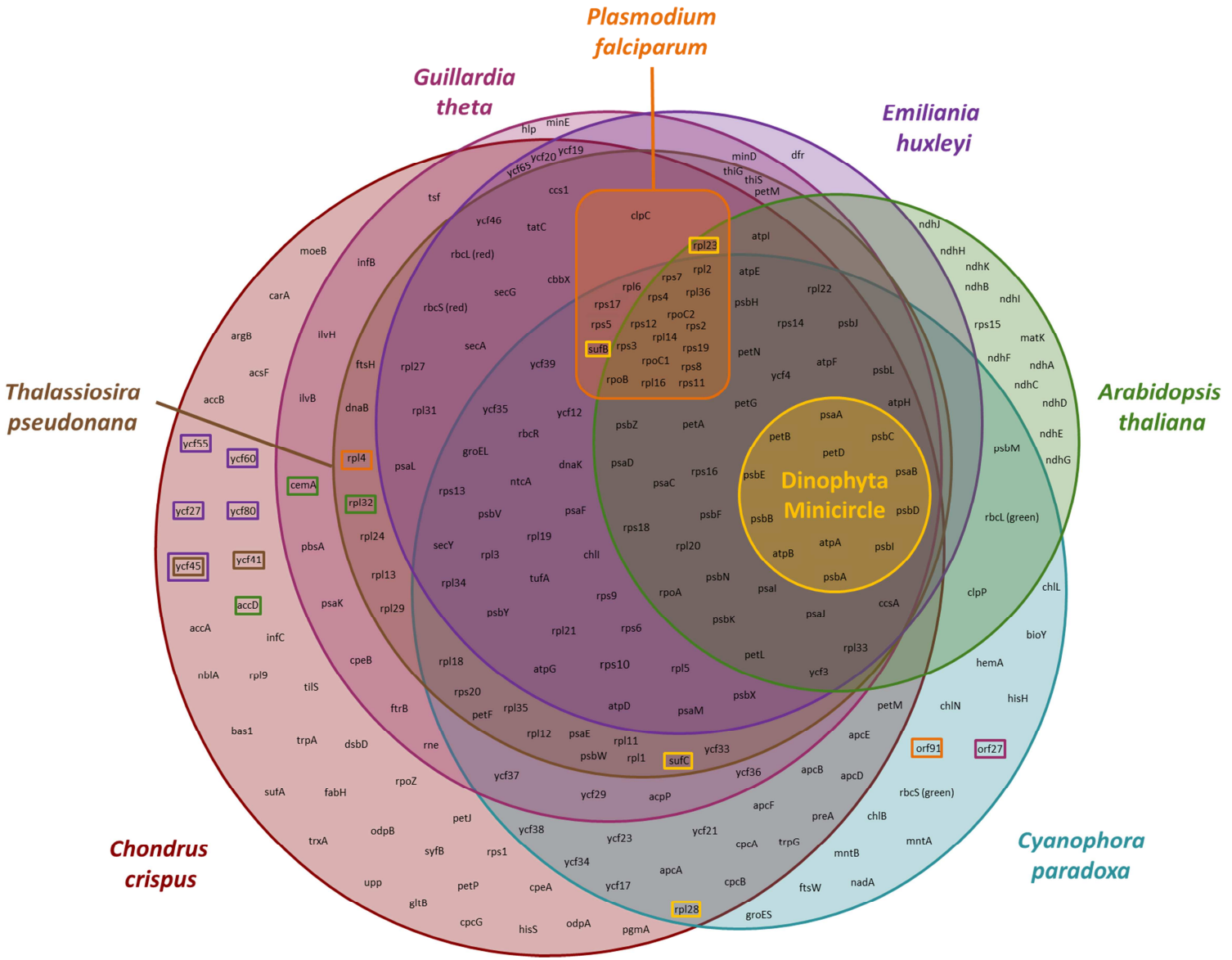


Figure 20: Venn diagram of plastid genomes in eight eukaryotes. Boxed genes are shared with the species of the equivalent color. ycf and orf symbolize hypothetical proteins. Only shared hypothetical proteins are shown.

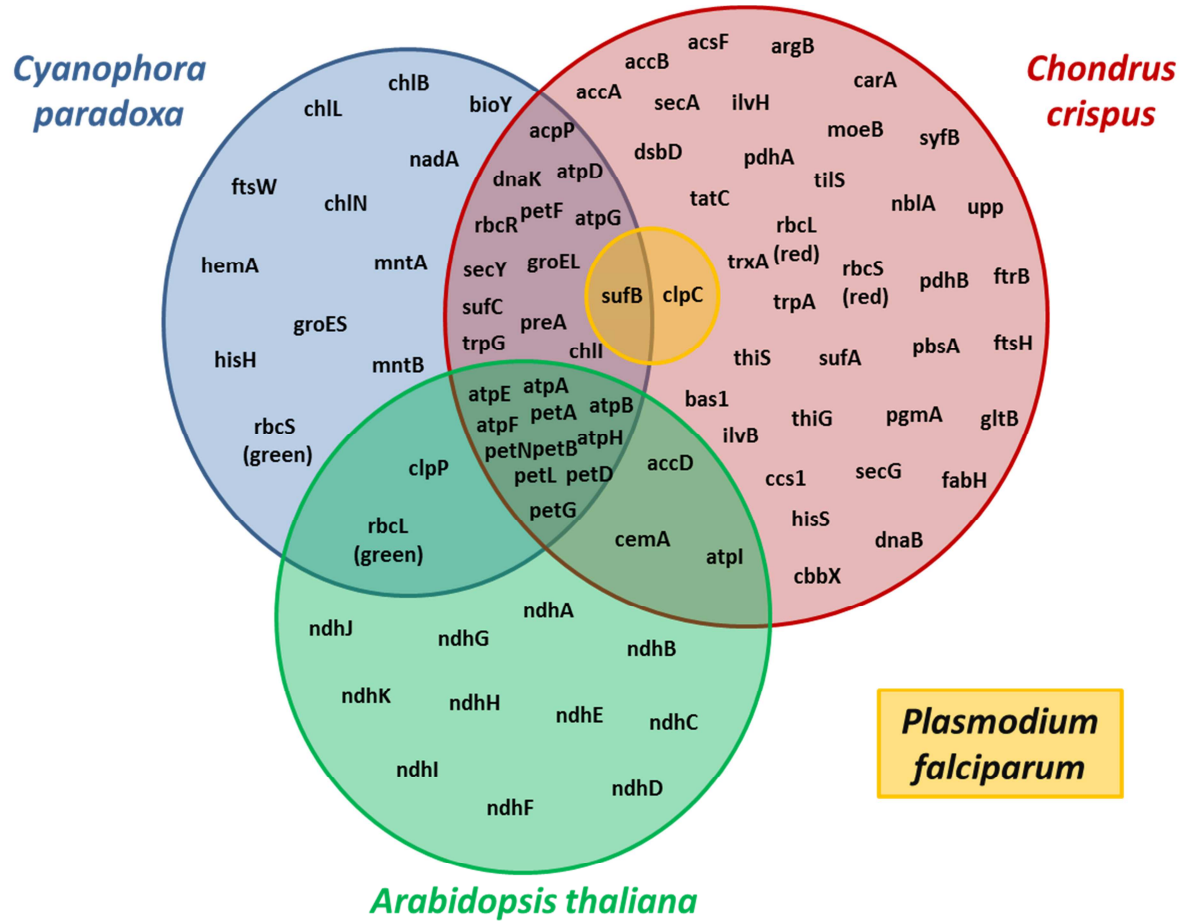


Figure 21: Venn diagram for the plastome of three species harboring a primary plastid (*C. paradoxa* blue, *C. crispus* red and *A. thaliana* green) and of the apicoplast genome of the malaria pathogen *P. falciparum* (yellow). The amount of plastidial genes is reduced to genes involved in metabolic pathways of the plastid: Biosynthesis, Transport, Cell processes and Energy metabolism (categories defined in (Sanchez Puerta et al. 2005).

In this analysis only two genes are left in the apicoplast of *P. falciparum* shared with the other plastomes: the chaperon *clpC* and the cysteine desulfurase *sufB*. These genes were checked via local blast analyses in the metagenome and transcriptome of *P. olseni.* In their homologs introns were found, thus documenting their nuclear localization (Table S5). In the next step, essential plastid genes were targeted, which are encoded on the plastome of red algae, but are nucleus-encoded in *P. falciparum* due to the fact that the apicoplast is a plastid of "red algal origin" (Chapter 1.2.2). To avoid paralogous genes, these genes were checked if they were absent in aplastidial alveolate species like the ciliate *T. thermophila* or the apicomplexan parasite *C. parvum* (Table S6). The local blast analysis revealed that four genes were fitting to these criteria: the carbamoyl phosphate synthase gene *carA* for pyrimidine and arginine biosynthesis, the β-ketoacyl-carrier protein synthase III gene *fabH* for fatty acid biosynthesis, the anthranilate synthase gene for tryptophane biosynthesis *trpG* and the soluble (2Fe-2S) ferrodoxin *petF*. Likewise, all these genes are present and nucleus-encoded in *P. olseni*; the

gene *fabH* excluded, which is missing in the dinoflagellate parasite (Table S5/Table S6). Based on the observation that these essential genes for plastid functions are encoded in the nucleus of *P. olseni* and due to the absence of the 16S rRNA, which represents the protein biosynthesis machinery (see above), I propose that this eukaryotic parasite harbors a plastid, but lost its plastome.

Presence of Nuclear-encoded Plastid-targeted Genes

The presence of a plastid implies the presence of essential metabolic pathways proceeding in the lumen of this organelle. An example for such a metabolic pathway is the extensively studied MEP pathway for isoprenoid biosynthesis (Grauvogel et al. 2007; Matsuzaki et al. 2008; Lohr et al. 2012). It was shown that *Plasmodium* does not survive if one protein of this pathway, the 1-deoxy-d-xylulose 5 phosphate (DXR), is inhibited by the herbicide fosmidomycin (Jomaa et al. 1999). In 2011, Yeh and DeRisi could rescue the fosmidomycin treated blood stage cells with the supplementation of isopentenyl pyrophosphate (IPP), the product of the MEP pathway and the precursor molecule for more complex isoprenoids. Interestingly, the permanent treatment of the cells with IPP resulted into plastid genome cured strains of *P. falciparum* (Yeh and DeRisi, 2011). This experiment showed that the apicoplast of this parasite is only maintained because the cell needs the plastidial production of the IPP in absence of the cytosolic MVA pathway for isoprenoid biosynthesis. In my analyses of apicomplexan and dinoflagellate key organisms (*C. velia*, *V. brassicaformis* and *P. olseni*) and also in the dinoflagellate reference species *P. minimum*, the cytosolic MVA pathway is missing, although the first gene acetoacetyl-CoA thiolase (*aact*) of this pathway could be detected. Thus, *P. olseni* could have retained its heterotrophic plastid in order to ensure the production of the isoprenoid precursor IPP. The presence of the MEP pathway genes with plastidial bipartite signal sequences in the genus *Perkinsus* was already the subject of several studies (Grauvogel et al. 2007; Matsuzaki et al. 2008). Likewise, in the *P. olseni* transcriptomic dataset of this study all seven MEP genes could be identified. Six of them (*dxs, dxr, mct, mecps, hds and hdr*) are in full length, whereas the *cmk* gene seems to be partial, but could be completed with the genomic sequence (Table 4).

Spliced leader (SL) trans-splicing of mRNA sequences are common in dinoflagellates. Therefore, 21-22 nt of non-coding RNA tandem repeats are transferred as a SL sequence to the N-terminus of mRNA of nuclear genes (Zhang et al. 2007). Thus, these pre-sequences are, apart from the characteristic polyA+ tails for the initial isolation of nuclear–encoded mRNA, a marker for the nuclear localization of genes. In six out of seven MEP genes these SL

sequences were identified in the transcriptomic dataset of *P. olseni* emphasizing their status as nuclear-encoded plastid genes (Figure 22).

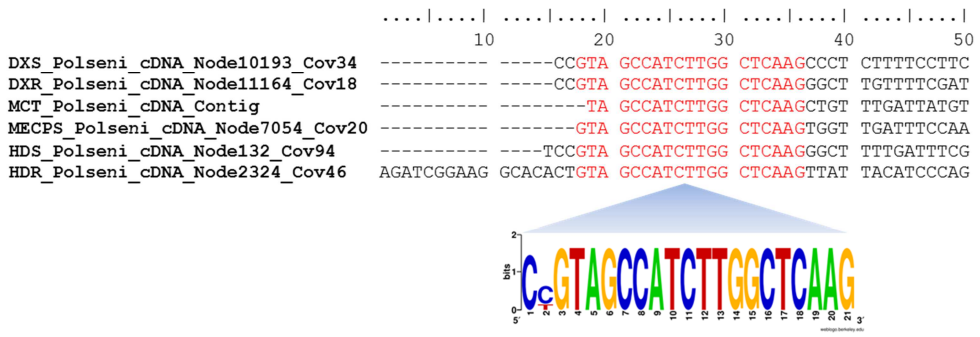


Figure 22: Alignment of the N-terminus of six MEP genes (DXS, DXR, MCT, MECPS, HDS, HDR). The dinoflagellate spliced leader sequence (SL) is highlighted in red. The SL sequence is summarized in a WebLogo (Crooks et al. 2004); http://weblogo.berkeley.edu/) graphic below.

The SL sequence of *P. olseni* is highly conserved and almost identical to the general SL sequence of dinoflagellates (Zhang et al. 2007) and to the already described SL sequence of *P. marinus* (Zhang et al. 2011). Additionally to the presence of a SL sequence, multiple introns were found in the MEP genes if compared to the genomic dataset of *P. olseni* (Table 4) confirming their nuclear localization. Furthermore, the genomic coverage values of MEP genes are comparable to the on average 140-fold nuclear coverage in *P. olseni* (Table 4).

Table 4: Summary of MEP protein analyses in *P. olseni*. The presence of full length cDNA and the dinoflagellate splice leader sequence (SL) is marked by a check mark, absence by a cross. Abbreviations: aa = amino acids, CS = cleavage sites predicted by [program]. The full length protein sequence of CMK is inferred by a combination of the genomic and transcriptomic data and is marked by an asterisk.

MEP proteins	DXS	DXR	MCT	CMK	MECPS	HDS	HDR
Full length cDNA/SL	✓	✓	✓	x	✓	✓	✓
Scaffolds/contigs [#]	10193	11164	252,253	8504	7054	132	2324
Genome coverage	118x	97x	191x	126x	128x	119x	118x
Protein length [aa]	849	534	464	381*	240	795	582
Introns	23	11	4	4	6	23	15
CS [SignalP]	32‖33	38‖39	25‖26	34‖35	30‖31	34‖35	43‖44
CS [TargetP]	100‖101	106‖107	81‖82	60‖61	93‖94	126‖127	102‖103
"FVAP" motif	FIQP	FVLG	FYSP	FVFP	(FVPL)	FLLP	FVSP

The next aim was to demonstrate that these MEP proteins contain a bipartite signal sequence for a proposed complex plastid. Dependent on the amount of membranes surrounding the plastid, primary plastids only contain a transit peptide to conquer two membranes, whereas secondary plastids contain a bipartite signal sequence consisting of a signal peptide and a transit peptide to cross three (Dinoflagellates) to four (Apicomplexa) surrounding membranes (Chapter 1.2.2). The reliable detection of these pre-sequences is bothered by their low conservation frequently resulting in contradictory results of the different prediction programs.
I generated a prediction pipeline in case of the MEP genes of *P. olseni* in order to detect their bipartite signal sequences for the import into a complex plastid. Previously MEP protein sequences were aligned with MEP protein sequences of bacteria (highest blastp hits), mostly cyanobacteria, which do not contain any signal pre-sequences, in order to estimate the length of a possible bipartite signal sequence (>100aa residues, Figure S9). The cleavage site of the signal peptide was predicted with the program SignalP v.3.0 truncating the sequence automatically to a maximum of 70 aa residues and using the method of a Hidden Markov Model (HMM; Nielsen et al. 1997; Nielsen, Krogh 1998) for cleavage site prediction in eukaryotes. The more or less well supported cleavage site predictions are summarized in Table 4. Furthermore, the conserved "FVAP" motif was identified up- and downstream of the SignalP predicted cleavage site as a marker for the starting point of transit peptides in dinoflagellates (Patron et al. 2005). Variants of the "FVAP" motif were found in all seven MEP proteins (Table 4), whereas in the MECPS protein this motif is not in close spatial vicinity to the signal peptide cleavage site unlike the motifs of the other MEP proteins (Figure S9). Using the "FVAP" motif as a transit peptide starting point, the rest of the N-terminal aa residues of MEP genes was utilized as an input for another prediction program for signal sequences, TargetP v.1.1 (Emanuelsson et al. 2000). In five out of seven MEP proteins (DXS, DXR, MCT, HDS and HDR) a plastidial localization was clearly detected. In the CMK and MECPS protein a mitochondrial localization was proposed. However, the predicted transit peptide cleavage site of every MEP protein was recorded (Table 4). Given the detected bipartite signal sequence, the transit peptide can be classified into class I, class II and even class III transit peptides in dinoflagellates (Patron et al. 2005). The class I transit peptides can be distinguished containing an additional hydrophobic transmembrane domain as a stop-transfer membrane anchor within the transit peptide. Transmembrane domains were calculated in the first >100 aa of MEP proteins using the program SOSUI (Hirokawa et al. 1998). This approach was combined with Kyte-Doolittle hydropathy plots of every MEP protein sequence (graphical revision in Figure S9) using the program ProtScale (Gasteiger et

al. 2005) implemented in the web server of the Bioinformatics Resource Portal ExPASy. Four out of seven MEP proteins were detected to contain class I transit peptides. Their protein sequences (in general the first 150 aa residues) were aligned with the MUSCLE algorithm in MEGA 5.0 (Tamura et al. 2011) and the Kyte-Doolittle plot for this alignment was recalculated using the program pepwindowall of EMBOSS (Rice et al. 2000) with a window size of 9 (Figure 23). Interestingly, the MUSCLE algorithm aligned the "FVAP" motifs of each MEP protein one below the other (red box, Figure 23) which reinforces the signal as the transit peptide starting point. The first and fourth position within this motif seems to be conserved (asterisks, Figure 23).

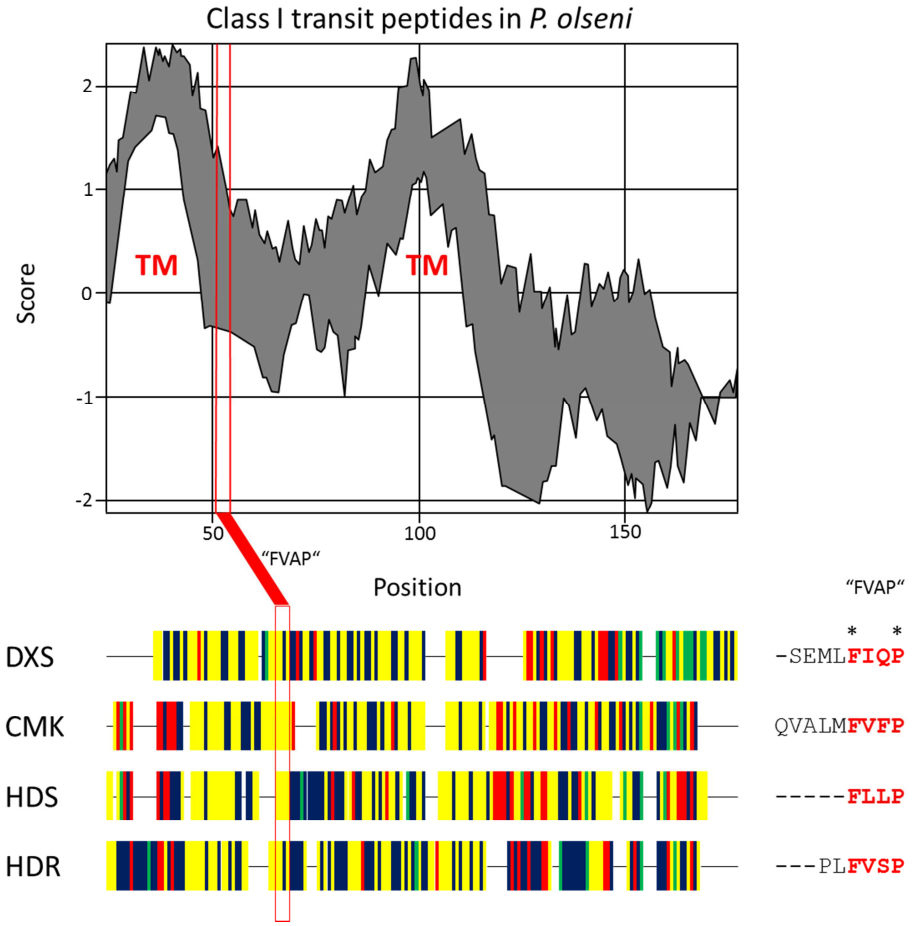


Figure 23: Alignment of the first 150 aa of the MEP protein sequences DXS, CMK, HDS and HDR showing a plastidial bipartite signal sequence with a class I transit peptide of dinoflagellates (Patron et al. 2005). The amino acids are classified into groups (yellow: hydrophobic, non-polar; blue: polar, hydrophilic; red: basic, positively charged; green: acidic, negatively charged). The position of the conserved 'FVAP' motif is highlighted in red. Conserved positions within this motif are marked by an asterisk. The range of the Kyte-Doolittle hydropathy analyses for the protein alignment is shown in grey in the plot above. The transmembrane domains (TM) are highlighted in red. The single analyses to each MEP protein are shown in Figure S9.

Transit peptides lacking an additional hydrophobic domain are designated as class II transit peptides (Patron et al. 2005). The Kyte-Doolittle hydropathy plots of the DXR and MECPS protein (Figure S9) are comparable showing only one transmembrane domain within the signal peptide. Thus these two proteins could be aligned and reanalyzed as class II transit peptides (Figure 24). In this case the "FVAP" motif is not vertical aligned, but the first two amino acids of this motif are identical (asterisks, Figure 24). The protein sequence of the MEP protein MCT has no explicit classification. In its hydropathy plot no transmembrane domain in the signal peptide is detectable, whereas such a domain is clearly present within the predicted transit peptide (Figure S9). The structure of the N-terminal MCT protein sequence resembles that of class II transit peptides, thus the transmembrane domain of the signal peptide seems to be shifted downstream.

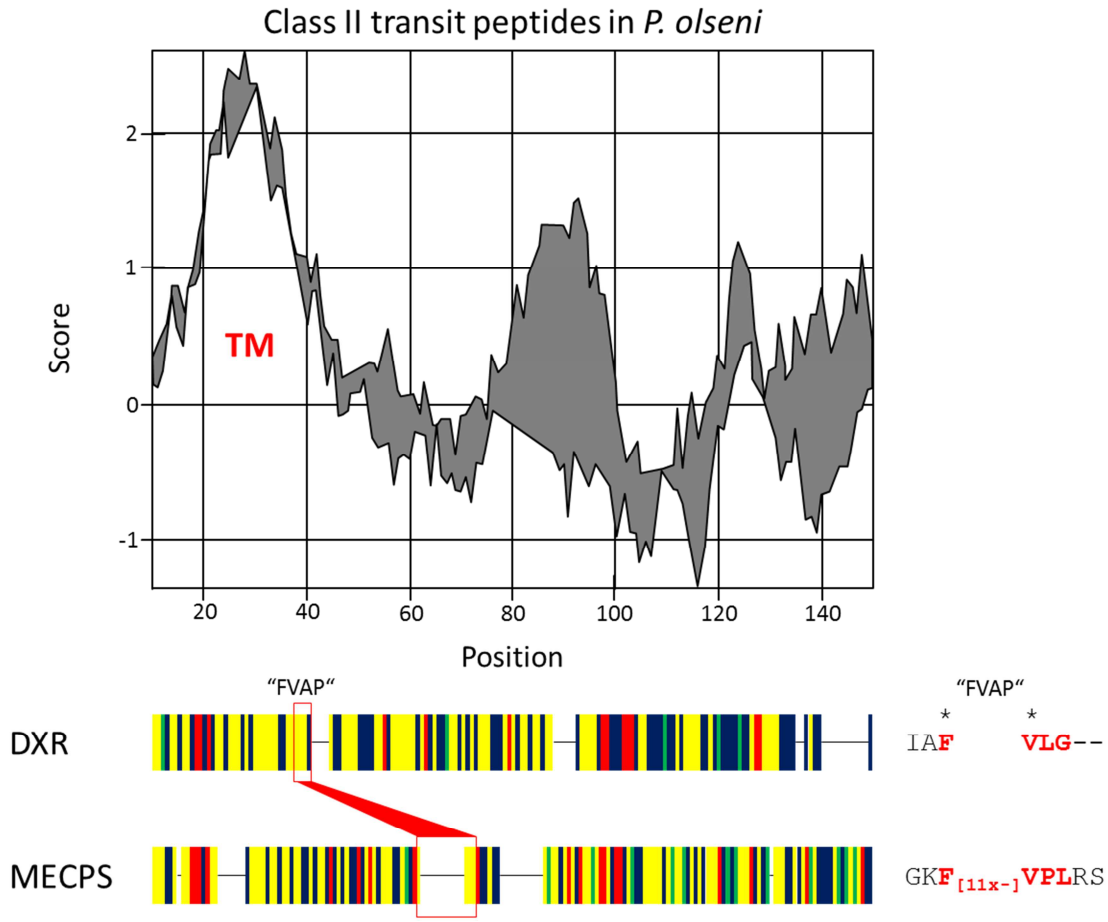


Figure 24: Alignment of the first 150 aa of the MEP protein sequences DXR and MECPS showing a plastidial bipartite signal sequence with a class II transit peptide of dinoflagellates (Patron et al. 2005). The amino acids are classified into groups (yellow: hydrophobic, non-polar; blue: polar, hydrophilic; red: basic, positively charged; green: acidic, negatively charged). The position of the conserved FVAP motif is highlighted in red. Conserved positions within this motif are marked by an asterisk. The range of the Kyte-Doolittle hydropathy analyses for the protein alignment is shown in grey in the plot above. The transmembrane domains (TM) are highlighted in red. The single analyses to each MEP protein are shown in Figure S9.

My analyses demonstrated the presence of plastid-targeted signal sequences, which refer to the working hypothesis that the MEP proteins are imported into a yet undiscovered complex plastid. All analyses indicated that *Perkinsus* spp. harbors this organelle, which differs from all other plastids described so far by the absence of a plastome. That is why the proposed term of the 'Perkinsuplast' by Grauvogel et al. (2007) is reasonable according to the use of 'Apicoplast' for the strongly reduced heterotrophic plastid in Apicomplexa.

Immunocytochemical Evidence for a Plastid in Perkinsus

Unfortunately, a final proof for the presence of a plastid structure via immunogold electron microscopy could not be accomplished due to methodical difficulties in the probe preparation step. However, two MEP gene sequences (*dxr* and *dxs*) of *P. marinus* were successfully amplified and cloned into the overexpression vector pQE80L. The corresponding polyclonal antibodies were produced and obtained by the company GeneCust immunizing four rabbits (2x DXR, 2x CMK). The control dot blots demonstrated no influence of the fixative to the binding behavior of the antibody. The best results were achieved with a combined fixative containing 1% formaldehyde and 0.2 % glutaraldehyde. The control western blots with both *P. marinus* (positive control) and *P. olseni* (sequenced organism) protein extracts were also successful, showing the best results for the antibody CMK_1 (RB3153; Figure S10). However, the results of the Western blots showed multiple bands in the total protein extracts which could be generated by: (1) a too high concentration of the antibody causing unspecific reactions; (2) the degradation of the CMK protein causing bands of lower molecular weight; (3) multimer forming of the target protein causing higher molecular weight bands.

Within their life cycle, the genus *Perkinsus* produces biflagellated zoospores, which have an important role in the distribution of this parasite and potentially in the infection of their crustacean hosts (Perkins 1976). These motile cells show a clear and condensed arrangement of cell organelles in comparison to other cell stages (Sunila et al. 2001), which makes them very attractive for the use in electron microscopy. After successful enrichment of the zoospore-stage in the *P. olseni* cell culture, the cells were fixed and prepared for immunogold electron microscopy. The micrographs showed spherical cells with large vacuoles containing a structure of high electron density (Figure S11 A). The antibody bound relatively unspecific with a slightly higher accumulation of gold particle spots in these electron dense structures (Figure S11 B). Unfortunately, the resolution of the micrographs is not that high, thus the only recognizable structures could be identified as the nucleus, the vacuoles and the electron dense structure within. Furthermore, the micrographs propably show the wrong cell stage.

Zoospores are elliptical cells characteristic for flagella and a conoid associated with micronemes (Coss et al. 2001; Sunila et al. 2001). None of these structures could be detected in the micrographs, whereas the numerous vacuoles containing the electron dense structures seem to be characteristic for trophozoites or pre-zoospores, respectively. The electron dense structures could represent in this case so-called 'vacuoplasts' which serve as internal storage for proteins in the cell. They possibly develop from droplets of the endoplasmatic reticulum (Sunila et al. 2001). The question arises what happened with all zoospores that were enriched in our cell culture. It is obvious that the immunogold method has to be improved for the visualization of zoospores of *P. olseni*. Therefore, further immunogold electron microscopic experiments are necessary that is due to the extensive and time-consuming cultivation challenging.

Reconstruction of Essential Plastid Functions in the 'Perkinsuplast'

Taken together, the maximum reduction of the plastome seems to be achieved in *Perkinsus* spp.: genes of essential plastid functions were transferred to the nucleus, whereas other genes of the plastome got lost. Via local blast analyses the essential plastid functions were summarized in a schematic overview of a predicted 'Perkinsuplast' in *P. olseni* according to the style of Gornik et al. (2015). The plastid itself is illustrated as an organelle surrounded by either three (continuous lines, dinoflagellate plastid) or four (additional dashed line, Teles-Grilo et al. 2007) surrounding membranes as the final evidence for its composition is missing (Figure 25).

Nuclear-encoded plastid-specific genes of the MEP pathway for isoprenoid precursor biosynthesis as well as the SUF pathway for iron-sulphur cluster assembly were identified (Table S5). Plastid-targeting sequences indicate the localization of these enzymes in the plastid as extensively studied for the MEP pathway in the previous subchapter and as shown by the detection of targeting sequences with the web-based prediction programs TargetP (Emanuelsson et al. 2000) and iPSORT (Bannai et al. 2002) for the SUF pathway (Table S5; Figure 25). Additionally, it was shown that a complete functional plastidial FASII pathway for fatty acid biosynthesis is missing, although the nuclear-encoded gene for the enzyme FabG was detected, but carrying a mitochondria-targeted pre-sequence (Table S5). By confirming the presence of the multienzyme polyketide synthase of the FASI pathway, it seems that the retention of this cytosolic pathway promoted the loss of the plastidial fatty acid pathway by replacing it (Table S5; Figure 25).

The tetrapyrrole biosynthesis pathway seems to be a hybrid of both, the mitochondrial C4 and the plastidial C5 pathway, similar to the composition in Apicomplexa (Koreny et al. 2011). Two scaffolds coding for the 5-aminolevulinic acid synthase (ALAS) are present. Unfortunately, the N-terminus of these enzymes is missing and thus the pre-sequence for a localization analysis. However, the sequence similarity to the mitochondrial ALAS in *Hematodinium* sp. (AKG25418.1, Gornik et al. 2015) is so high that its localization could be predicted likewise in the mitochondrion (Figure 25). The next steps of tetrapyrrole synthesis seem to be performed by enzymes of the C5 pathway in the plastid. The genes coding for HemB, HemC and HemE were identified carrying plastid-targeting signals (Table S5), whereas the intermediate enzyme HemD could not be found, perhaps due to its loss or its highly diverse sequence, which often occurs in fast evolving parasites. The same could be applied to the missing enzyme HemF and HemH. In contrast, the gene coding for the mitochondrial protoporphyrinogen oxidase (HemY/G) was detected indicating that the remaining steps of tetrapyyrole synthesis have to take place again in mitochondria (Table S5; Figure 25).

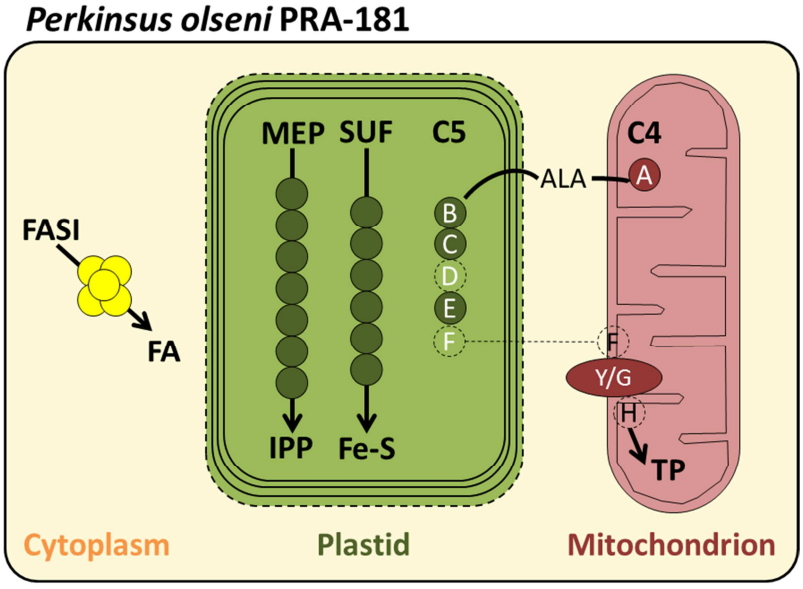


Figure 25: Summary of essential myzozoan plastid functions (Lim, McFadden 2010; Gornik et al. 2015) and their localization in *P. olseni* (strain PRA-181). The graphic is based on blast and prediction analyses in the transcriptomic as well as genomic datasets of *P. olseni* (Table S5). Enzymes are represented by colored circles; missing enzymes are illustrated in transparent dashed circles. Abbreviations: FA = fatty acid; FASI = fatty acid synthesis pathway I; MEP = 2-C-methyl-D-erythritol 4-phosphate pathway; IPP = isopentenyl pyrophosphate; SUF = iron-sulphur cluster biosynthesis pathway; Fe-S = iron-sulphur cluster; C4/C5 = tetrapyrrole biosynthesis pathways; TP = tetrapyrrole.

3.1.5 The 'Red' Complex Plastid of *C. velia* and a 'Green' Footprint

The genomes of protists have a mosaic composition significantly marked by horizontal gene transfer either of endosymbiotic or of xenologous origin. For example, xenologous replacements of nuclear-encoded plastid genes by the bacterial order Chlamydiales were detected for the MEP enzymes HDS and HDR in this thesis (Chapter 0.43.1.2; Figure 14; Petersen et al. 2014) and were also observed for two additional MEP enzymes MCT and CMK. In the photoautotrophic alga *C. velia* it was demonstrated that this species harbor a complex plastid of red algal origin (Janouskovec et al. 2010), which could be accompanied with the described apicoplast monophyly including *C. velia* in this study (Chapter 0.43.1.2). Surprisingly, in a broad phylogenomic study based on expressed sequenced tags (ESTs) of *C. velia* a large number of genes of green algal ancestry were found (Woehle et al. 2011). However, these published data were re-analyzed by (Burki et al. 2012), because they found indications of contaminations by land plants in these ESTs. Nevertheless, although the number of green signals in algal-derived genes could be reduced by a new analytical pipieline, green 'footprints' still remained (Burki et al. 2012).

These observations are compatible to one discovery in the new established datasets for *C. velia* in this work regarding one protein of the plastidial SUF pathway: the component SufB of the Fe-S cluster scaffold protein complex that assembles Fe-S clusters (Lill 2009; Chapter 1.5.2). Performing a blastx search with the nuclear-encoded plastid-targeted *sufB* gene of *C. velia*, exclusively hits to cyanobacterial sequences were obtained, indicating that the SUF pathway, which was once acquired by the first cyanobacterial endosymbiont, outlived further endosymbioses during the evolutionary history of this complex alga. Surprisingly, under closer scrutiny of the intron distribution based on a protein alignment including representatives of green plastids, *C. velia* shares seven out of 10 intron positions with green algae. Notably, six out of these seven shared intron positions are identical to positions of the chlorophyte *Chlamydomonas reinhardtii* (Figure S12). Thus, *C. velia* seems to have acquired the *sufB* gene in a xenologous context from a green alga, most likely from a precursor of *C. reinhardtii*. This would also explain the discrepancy of nuclear-encoded SufB in *C. velia* and the plastid-encoded SufB in the other parasitic apicomplexan species and even the closest relative *V. brassicaformis*.

A phylogenetic analysis of the SufB protein was performed in order to investigate the origin of the *C. velia sufB* gene (Figure S13). Cyanobacteria were chosen to form the basis of this tree. The phylogeny shows a clear arrangement of groups containing plastid-encoded SufB sequences (CASH, red algae and *C. paradoxa*) and groups containing nuclear-encoded SufB

sequences (‘green lineages’ with primary and secondary plastids: Viridiplantae, Euglenozoa, Rhizaria). Although the tree is not well resolved due to the fast evolving SufB sequences of the apicoplast, a ‘green origin’ of the also nuclear-encoded SufB protein of *C. velia* cannot be excluded: *C. velia* clusters together with the dinoflagellate sequences, though with only weak statistical support, but forming the sister group to a big cluster consisting of the other nuclear-encoded ‘green’ SufB sequences indicating a ‘green affinity’ of this protein in the chromerid.

3.2 Mitochondrial Evolution in Alveolata

"[...] *it is obvious that the evolution of mitochondrial genomes is reductive* […]"

(Flegontov et al. 2015)

3.2.1 *Chromera velia* and the Smallest Mitochondrial Genome Found in Algae

Examining the mitochondrial genome of *C. velia* only the mitochondrial cytochrome c oxidase subunit I gene (*coxI*) within the new high qualitative genomic dataset was found representing the complex IV of the respiratory electron transport chain in *C. velia*. This analysis was conducted via local tblastn searches with the reference sequences of *P. falciparum* and *O. marina*. The result led to the hypothesis that this alga possesses a mitochondrial genome consisting of only one single protein-coding *coxI* gene. Moreover, the question arose if this genome has a circular conformation or if it is linear and therefore comparable to the apicomplexan representatives.

In order to answer the question of conformation two different approaches in the laboratory were conducted in parallel. A) In the case of a genomic circularity inverse PCR with nested primer pairs within the coding region of the *coxI* gene was performed. B) On the other hand tailing experiments to determine the genomic ends of a hypothetical linear mitochondrion were accomplished following the instructions of Hikosaka et al. (2010). Therefore, tailed genomic DNA of *C. velia* with a polyC-tailing was generated and the DNA was preheated and treated with 3% formamide to ensure its linear conformation.

It was possible to amplify mitochondrial fragments via the inverse PCR technique (Figure S14) with partly different terminal sequences. This "bridging" PCR result would infer both: more than one mitochondrial minicircle with the *coxI* gene and different non-coding DNA regions; or a temporarily circularized or concatenated genome connected by terminal inverted repeats of a linear mitochondrial genome molecule. Since the tailing experiment was not successful the experimental data would support a circular structure of the mitochondrial genome of *C. velia*. The experimental sequence data (iPCR) were merged with the Illumina sequenced nodes to generate a valid elongated contig sequence that should represent the mitochondrial genome of *C. velia* (Figure S14; Petersen et al. 2014). The *coxI* gene that was found via local blast analyses in the Illumina dataset was only partial, but it was possible to reconstruct the full length protein-coding gene with the inverse PCR experiment (Figure S14). In summary five Illumina nodes (Node186709, Node235449, Node461926, Node2236 and Node119218) could be identified with a coverage range of 98x-695x describing the

mitochondrial genome of *C. velia*. The contig sequence produced a molecule with an expected size of 1971 bp (KC899110) which was already published in our paper about *C. velia* and the Rhodoplex hypothesis (Figure S14; Petersen et al. 2014). Further investigations on the iPCR products in combination with the sequence data of Illumina revealed that the sequences could be circularized to two different mitochondrial genomic molecules, which vary in their sizes. The combination of iPCR2 and iPCR3 (Figure S14) produces a molecule with a length of 2158bp disregarding that this molecule is not homologous to the N-terminus of the Illumina_1 node (position 208 – 260 in the alignment). However, this partial sequence of Illumina_1 is homologous to the sequence of iPCR4, which can be also circularized to the smaller mitochondrial genome molecule of 2072bp. Both results indicate that there exist different mitochondrial genomic minicircles in *C. velia* of around 2kB, which differ in their sizes by various non-coding DNA regions. Therefore, the sequence of Illumina_1 could represent in combination with iPCR4 one version of non-coding DNA region and both independent iPCR experiments 2 and 3 could constitute the second version of non-coding DNA region in such a mitochondrial minicircle.

In 2015, Flegontov et al. described numerous short linear mitochondrial molecules with a single conserved *coxI* region, a highly diverse *cox3* gene as well as numerous rRNA fragments in *C. velia*. Their data rely on mitochondrial DNA-enriched fractions of *C. velia* genomic DNA. They mentioned that this highly diverse *cox3* sequence is missing in our dataset. However, their putative *cox3* sequence was inferred by the percentage of its hydrophobic residues, number of predicted transmembrane domains (which is comparable to *Plasmodium*) and very short conserved myzozoan *cox3* motifs. Furthermore, they pointed out that in dinoflagellates the *cox3* sequence is likewise divergent (Flegontov et al. 2015). In order to verify the presence of a divergent *cox3* sequence in our genomic dataset, local blast analysis were performed with the proposed protein coding sequence of *cox3* by Flegontov et al.. Two highly similar hits for this sequence were identified: Node133522 (69x coverage) and Node132247 (294x coverage). Both nodes could be fused to a contig sequence with a length of 1302 bp. Subsequently, blastx and blastn searches against the NCBI database were conducted with these two newly acquired mitochondrial nodes. The blastx search with Node132247 identified no significant similar sequences in the NCBI database. Interestingly, the blastx search with the other Node133522 resulted only in a hit with our published amino acid sequence of *coxI*. Actually, the first 60 nt of the reverse complement of this node overlap with a sequence region in the middle of our *coxI* coding sequence. Blastn searches with both nodes just resulted in highly similar hits with the clones of the Czech working group and

genomic hits of low support for other eukaryotic non coding sequences. The new supposed mitochondrial sequence was aligned to the previous experimental alignment (Figure S14), but apart from the homologous reverse complement sequence region to our *coxI* sequence, no similarity to the sequences in the alignment were found (data not shown). Accordingly, cox3 homologs are missing in the *C. velia* genome, but its functional role in the mitochondrion might be conducted by an non-homologous equivalent.

However, I agree with this Czech working group that there exists probably not only one version of the mitochondrial genome in this photosynthetic alga. This hypothesis is supported by the appearance of different iPCR sequences alongside with the conserved protein-coding region of *coxI*. The possibility of a concatenated linear or temporary circular molecule by terminal inverted repeats cannot be excluded, although, it was not possible to detect such repeating sequences in the available experimental data. In any case, the mitochondrial genome of *C. velia* represents the most reduced molecule found in algae.

3.2.2 Comparative Mitochondrial Genomics in Alveolata

Based on the discovery of the smallest mitochondrial genome in the early-branching apicomplexan species *C. velia* I was interested in the structure of other early-branching myzozoan mitochondrial genomes, e.g. *V. brassicaformis* and *P. olseni*, to get new insights into their evolutionary behavior.

The Mitochondrial Genome of V. brassicaformis

In a recent publication, some characteristics of the *Vitrella* mitochondrial genome have already been described: (1) the mitochondrion contains three protein-coding genes (*coxI*, *cob* and a divergent *cox3*), (2) the *coxI* and *cob* genes are fused, (3) extensive rRNA fragmentation was detected, (4) no RNA editing like in dinoflagellate mitochondria was found (Flegontov et al. 2015). In my genomic dataset of *V. brassicaformis* the tblastn search with queries of *C. velia*, *T. parva*, *O. marina*, *P. minimum* and *Pfiesteria piscicida* revealed only two independent mitochondrial protein-coding genes, *coxI* and *cob*, which were not fused. This example shows that automatic assemblies of mitochondrial genomes based on the Illumina sequence data are in general complicated. A central assembly function is the choice of the k-mer size. This value determines to what extent possible subsequences of the length k overlap in reads that can be assembled to contigs (Zerbino, Birney 2008). Smaller k-mer sizes increase the amount of overlapping reads, but can also result in false assemblies by repetitive and random similar sequences. Larger k-mer sizes reduce these false assemblies for small

repeat regions, but also lead to smaller contigs as a larger k-mer can only overlap with a restricted number of other k-mers. Thus, the appropriate k-mer has to be chosen very careful. The k-mer size chosen for the genomic dataset of *V. brassicaformis* was 41 bp, perhaps too large to detect a fusion between the *coxI* and *cob* reads.

In an independent approach putative contigs of the mitochondrial genome were identified based on their coverage. Because of their relative abundance in the cell, mitochondrial sequences have in general higher coverage values than plastid and mainly nuclear sequences. The coverage values of the detected *cob* and *coxI* sequences are 557x and 535x, respectively. In order to get a vague idea of the size and structure of the mitochondrial genome in *V. brassicaformis* a file was prepared containing 158 assembled genomic nodes with coverage values higher than 200x to find more putative mitochondrial sequences. By means of blastx and blastn searches in the NCBI web server, 68 of these nodes were assigned to chloroplastidic origin, 16 out of 158 were identified as putative mitochondrial sequences. These 16 sequences, as well as the entire 158 sequences dataset, were newly assembled with an implementation in Geneious v9.0.4 (http://www.geneious.com, Kearse et al. 2012). Six sequences of the reduced dataset (16 sequences) could be combined to a consensus sequence of 2926 bp (data not shown). The assembly of the 16 sequences was compared with the assembly of the 158 sequences and another node (Node_238962) was found overlapping with a node of a part of the *cob* sequence (Node_107822). The sequence was integrated into the reduced mitochondrial dataset and a *de novo* assembly of the sequences was accomplished. Actually, the newly added sequence was assembled to the other six mitochondrial sequences generating a consensus sequence of 3209 bp (Figure 26). Based on this consensus sequence a *cob-coxI* fusion is supported highlighting that a more complex assembly algorithm integrating also the coverage value information in combination with varying k-mers would be more accurate.

The translated consensus sequence was compared to the published amino acid sequence of the *cob-coxI* fusion by Flegontov et al. (2015) that resulted in an alignment with a 97% query coverage (1054 aa positives to our 1069 aa consensus sequence, Figure 27).

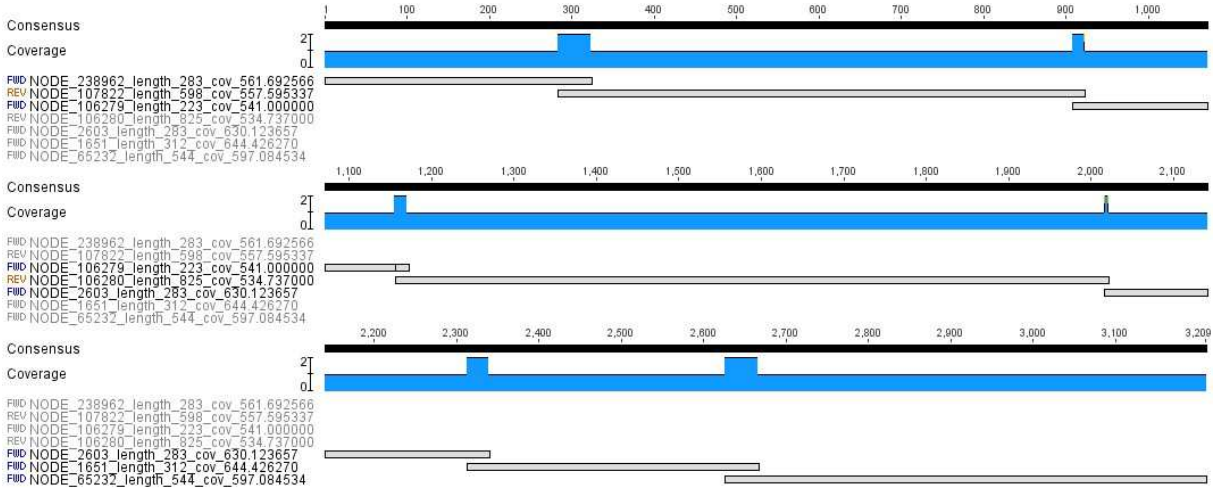


Figure 26: Assembly of putative mitochondrial sequences in the genome of *Vitrella brassicaformis*. Seven Nodes overlap to a consensus sequence of 3209 bp. The consensus sequence is shown as a black box with a ruler. The different overlapping nodes are presented in framed grey boxes whereas the coverage of the consensus sequence is highlighted in a blue box. Denoted are the Node identifiers with length and coverage values as well as their directions of the reading frame depicted in blue (forward, FWD) and brown (reverse, REV). Graphical visualization of the assembly by Geneious v9.0.4 (http://www.geneious.com, (Kearse et al. 2012).

Figure 27: Alignment of the 1069 aa mt genome sequence of *V. brassicaformis* and the 1096 aa sequence inferred by Flegontov et al. (2015). Our newly established mitochondrial consensus sequence (2.) is compared to the already published mitochondrial query of the *cob-coxI* fusion (1.). Additionally, the identity of both sequences is plotted in green, while a consensus sequence of the alignment is shown with a ruler above. Graphical visualization of the alignment by Geneious v9.0.4 (http://www.geneious.com, Kearse et al. 2012).

Thus, similar results for the *cob-coxI* fusion protein were obtained comparable to the Czech working group, although their COXI sequence seems to have a 13 aa insertion starting from alignment position 696. After blastp searches within the NCBI database, this insertion emerges as a real biological conserved region also found in other apicomplexan COXI

sequences. In the end, the k-mer size of the overlap between Node_106280 and Node_2603 (Figure 26) was chosen too small resulting in a false assembly of these two sequences to one contig. However, the 'gap sequence' could not be found in our genomic dataset of *V. brassicaformis* via local tblastn searches with the protein sequence of Flegontov et al. (2015). For his reason, additional PCR experiments in the N-terminus of the *Vitrella coxI* sequence has to be accomplished to deduce the full length of the *cob-coxI* fusion protein. This example highlights again the difficulty of the *de novo* assembly of the mitochondrial genome from Illumina genomic datasets by setting the wrong k-mer size. In general, larger k-mer sizes would be better suited for the *Vitrella* genome in order to reduce such false assemblies, also due to its bacterial contaminants. On the other hand, without the smaller k-mer sizes applied to the coverage filtered sequences, the other mitochondrial contigs would not have been found.

As in *C. velia*, the mitochondrial genome of *V. brassicaformis* should include a highly divergent *cox3* gene (Flegontov et al. 2015). Local tblastn searches were performed with appropriate *cox3* query sequences (*O. marina*, *S. minutum*, *P. falciparum*, *T. gondii*) and with the putative *cox3* protein sequence of Flegontov et al. (2015). Results were only obtained for the blast search with the putative *cox3* sequence of *Vitrella* (Node_66145) which has no similarities to any sequence of the NCBI protein database. A blastn search with this node resulted in the detection of mitochondrial sequences, but only in regions located in non-coding sequence areas. Comparable to former conclusions about the putative *cox3* sequence of *C. velia* in subchapter 0.43.2.1, the genome data do not support the claim of a highly divergent COX3 protein in *V. brassicaformis* (Flegontov et al. 2015). Furthermore, the putative cox3 sequences of Chromera and Vitrella are in no case homologous, although they are close relatives. It is reasonable that an ancient system like the mitochondrial oxidative phosphorylation is highly conserved and can thus be detected in the genome. Even in highly derived and evolved species, like the parasitic apicomplexans and dinoflagellates, the *cox3* sequences are as much conserved as detectable via conventional NCBI blast searches. For this reason, I assume that the mitochondrial genome in *V. brassicaformis* contains only two protein-coding genes (*cob* and *coxI*), which are fused to one open reading frame with a size of > 3000 bp (Table 5). In comparison to its chromerid relative *C. velia*, it has also experienced a strong reduction in its genomic content, but because of the presence of two protein-coding genes fused together, it is slightly larger than the smallest algal mitochondrial genome found in *C. velia*.

The question of the mitochondrion conformation in *V. brassicaformis* could not be clearly answered and inverse PCR experiments starting in the coding regions of *coxI* only resulted in fuzzy PCR bands. This outcome suggests a linear conformation as previously shown for parasitic Apicomplexa (Table 5).

The Mitochondrial Genome of P. olseni

The mitochondrion of *Perkinsus* spp. is poorly characterized, but intensively studied concerning unique frameshifts in their mitochondrial transcripts (Zhang et al. 2011). It was shown for *P. marinus* that the mitochondrial *coxI* genomic sequence and its mRNA sequence are identical, but could not be directly translated into its protein sequence (Masuda et al. 2010). At every AGG and CCC codon one translational frameshift was observed whose mechanism is still unknown. So far, only the *coxI* and *cob* gene was described to be located on the mitochondrial genome of *P. marinus* and *Perkinsus chesapeaki* (Masuda et al. 2010; Zhang et al. 2011).

Both mitochondrial genes were detected via local tblastn analyses in the genomic dataset of *P. olseni*: the *cob* gene on one single node (Node_104) and the *coxI* gene on two nodes (Node_3 and Node_594) which could be merged into one single contig. In terms of a possible *cob-coxI* fusion like in *V. brassicaformis*, a bridging PCR approach was performed with primer pairs in the coding regions of *cob* and *coxI* directed to each other. This approach was not successful indicating two plausible suggestions: 1) both genes cannot be connected by simple PCR methods, because the mitochondrial genome is expanded like in *S. minutum* (Shoguchi et al. 2015; Table 5) and too much bases separate the two open reading frames; 2) both genes are located independently on separate mitochondrial minicircles according to plastid minicircles in dinoflagellates. In order to test the second hypothesis, inverse PCR experiments were accomplished with primers starting in the coding regions of *coxI* and *cob*, but leading in opposite directions. In both cases, the iPCR experiments only led to PCR artifacts, which gave no indication about a possible circularity of the mitochondrial genome.

Furthermore, I generated again a genomic dataset of *P. olseni* restricted to comparable high mitochondrial coverage values. Based on the coverage values of *cob* (12861x) and *coxI* (15648x, 16025x), this dataset incorporated 203 sequences with coverage values >10,000x. Via blastx and blastn searches this dataset could be reduced to 31 putative mitochondrial sequences. Both datasets, the complete high coverage dataset and the reduced putative mitochondrial dataset, were *de novo* assembled with Geneious v9.0.4 (http://www.geneious.com, Kearse et al. 2012). Two separate contig sequences could be

generated incorporating the coding sequence for *cob* and *coxI*, respectively. A contig sequence of 11 nodes and a size of 7,284bp could be generated with the *cob* sequence. Only three nodes could be merged to a 2,785 bp contig describing the mitochondrial sequence around the *coxI* gene. Taken together the mitochondrion of *P. olseni* must have either a size of >10,000 bp, when both genes are located on one large mitochondrial molecule. On the other hand both genes could be located on different large independent molecules. The mitochondrial genome/s in *P.olseni* could have a linear conformation as the iPCR experiments were not successful.

Alveolate Mitochondrial Genomes

In comparison to other alveolate mitochondrial genomes all three early-branching myzozoan species have experienced a strong structural and genomic reduction in their mitochondria. Interestingly, the sister group to the Myzozoa, the colponemids (represented by *Acavomonas peruviana* in Table 5) are comparable to the original alveolate mitochondrion described in ciliates (exemplified by *T. thermophila* in Table 5), thus a strong selective pressure towards a reduction of the protein-coding genomic content must predominate in myzozoan mitochondria. From the ~45 protein-coding genes in original alveolate mitochondria, there remain only three protein-coding genes (*coxI*, *cob* and *cox3*) to the point of two protein-coding genes (*O. marina*: *coxI* and *cob-cox3* fusion) or even one protein-coding gene (*V. brassicaformis*: *cob-coxI* fusion; *C. velia*: *coxI*). This extreme reduction of the gene content would be expected for parasitic organisms, which experience the highest adaptation level to their environment and hosts. They are not completely reliant on their own metabolisms by scavenging important metabolites. In contrast, I found the most simplified mitochondrial genomes in two photoautotrophic algae and close relatives of the parasitic Apicomplexa (*V. brassicaformis* and *C. velia*).

In the case of the dinoflagellates (e.g. *S. minutum* in Table 5) the size of the mitochondrial genomes is strongly expanded by non-coding sequence regions whose function is not known (Shoguchi et al. 2015). Dinoflagellates are also renowned for their immense nuclear genome sizes and their collections of genes and non-coding sequences (Lin 2011; Wisecaver, Hackett 2011). In comparison the parasite *P. olseni* could have a similar expanded mitochondrial genome if both protein-coding genes *cob* and *coxI* were encoded on one large DNA-string (>10,000bp); likewise in *C. cohnii* and *O. marina*, whose final mitochondrial genome size could not be demonstrated until now. Excluded from these comparative analyses are the ‘dinotoms' like *D. baltica* and *K. foliaceum* emanated from a tertiary endosymbiosis event

with a stramenopile organism. Their mitochondrial genomes are very complex and comparable in size (~30 kb), content (>30 proteins) and form (putative circular molecule) to those of diatoms (Imanian et al. 2012).

Moreover, circular molecules in the alveolate superensemble seem to be a rarity. The discovery of varying non-coding regions by the iPCR experiments alongside with the *coxI* sequence in *C. velia* would support many versions of genomic minicircles harboring the same *coxI* gene, but differ in their non-coding sequences. However, the suggested circularity in *Chromera* could also be explained by a concatenated linear form where the protein-coding gene is surrounded by different non-coding spacer sequences.

Table 5: Comparison of mitochondrial genomes in the superensemble Alveolata. Mitochondrial characteristics of the alveolate species *T. thermophila* (Brunk 2003), *A. peruviania* (Tikhonenkov et al. 2014), *O. marina* (Slamovits et al. 2007), *C. cohnii* (Norman, Gray 2001), *S. minutum* (Shoguchi et al. 2015), *B. bovis* (Hikosaka et al. 2010), *E. tenella* (Hikosaka et al. 2010) and *P. falciparum* (Feagin et al. 1997; Feagin et al. 2012) were extracted from the literature. Newly established data from this study for *P. olseni*, *C. velia* and *V. brassicaformis* were added.

Group	Organism	size (bp)	organization	proteins	fusion
Ciliata	***T. thermophila***	47,577	linear	45	X
Colponemida	***A. peruviana***	51,109	linear	46	X
Dinoflagellata	***P. olseni***	>2,800, >7,300	linear?	2 (*coxI*, *cob*)	X
	O.marina	?	multipartite linear	3 (*coxI, cob, cox3*)	✓ (*cob-cox3*)
	C. cohnii	?	multipartite linear	3 (*coxI, cob, cox3*)	X
	S. minutum	~326,000	?	3 (*coxI, cob, cox3*)	X
Apicomplexa	***C. velia***	~2,100	minicircle?	1 (*coxI*)	X
	V. brassicaformis	>3,000	linear?	2 (*coxI*, *cob*)	✓ (*cob-coxI*)
	B. bovis	6,005	linear	3 (*coxI, cob, cox3*)	X
	T. parva	5,895	linear	3 (*coxI, cob, cox3*)	X
	E. tenella	6,213	concatenated linear	3 (*coxI, cob, cox3*)	X
	P. falciparum	5,967	concatenated linear	3 (*coxI, cob, cox3*)	X

Phylogeny of the coxI-gene

The protein-coding gene *coxI* seems to be the core gene, which is found ubiquitous in all analyzed alveolate mitochondria and which is the last gene retained in the mitochondrial genome of *C. velia*. Its phylogenetic analysis (Figure 28) revealed explicit variations from the species tree (Figure 13). Both chromerid species are found in the dinoflagellate cluster: the *coxI* sequence of the *cob-coxI* fusion protein in *V. brassicaformis* is related to the early-branching dinoflagellate *O. marina* (60% BS) reflecting their similar mitochondrial genomic structure. Both species harbor a mitochondrial genome which encodes in each case one fusion protein carrying the sequence information for the COB protein and for *cox3* and *coxI*, respectively. However, *C. velia* clusters only weakly supported together with the *Perkinsus* genus and forms the deepest branch of the monophyletic myzozoan group (100% BS) excluding the dinoflagellate species *Kryptoperidinium* and *Lingulodinium*. The separate positioning of the *Perkinsus* genus in *coxI* and *cob* phylogenies gave already reason to the suggestion of an independent lineage designated as "Perkinsozoa" (Zhang et al. 2011). Likewise in this tree (Figure 28) the genus *Perkinsus* builds a separate cluster to all other dinoflagellate species, but the position is not well supported either. Possibly, *Chromera* and *Perkinsus* cluster together, because the coxI gene in the parasitic dinoflagellate is encoded on an independent mitochondrial genomic molecule alongside with another separate mitochondrial genomic molecule harboring only the protein-coding gene *cob*.

Surprisingly, as a sister to the myzozoan group, the apicomplexan parasite *C. parvum*, the colponemids (*Acavomonas peruviana* and *Colponema vietnamica*) and the two remaining dinoflagellate species (*K. foliaceum* and *L. polyedrum*) cluster together with high statistical support of 79% BS. Despite their close mitochondrial relationship to diatoms, the tertiary endosymbiotic dinoflagellate *K. foliaceum* has a monophyletic relationship (91% BS) to the secondary endosymbiotic dinoflagellate *L. polyedrum* and the other tertiary endosymbiotic dinoflagellate *D. baltica* clusters directly within all other dinoflagellate species.

The COXI sequences of the ciliates were observed to be very divergent in the alignment describing the heterotachy between the two ciliate subgroups (84% BS and 100% BS) in the tree. Perhaps the slower evolving COXI sequence represents the original orthologous COXI, whereas the other ciliate sequences could be the result of paralogy.

However, single gene analyses has to be interpreted with caution as in the case of the coxI tree they often do not allow reliable conclusions about general evolutionary trends in the respective cell compartment.

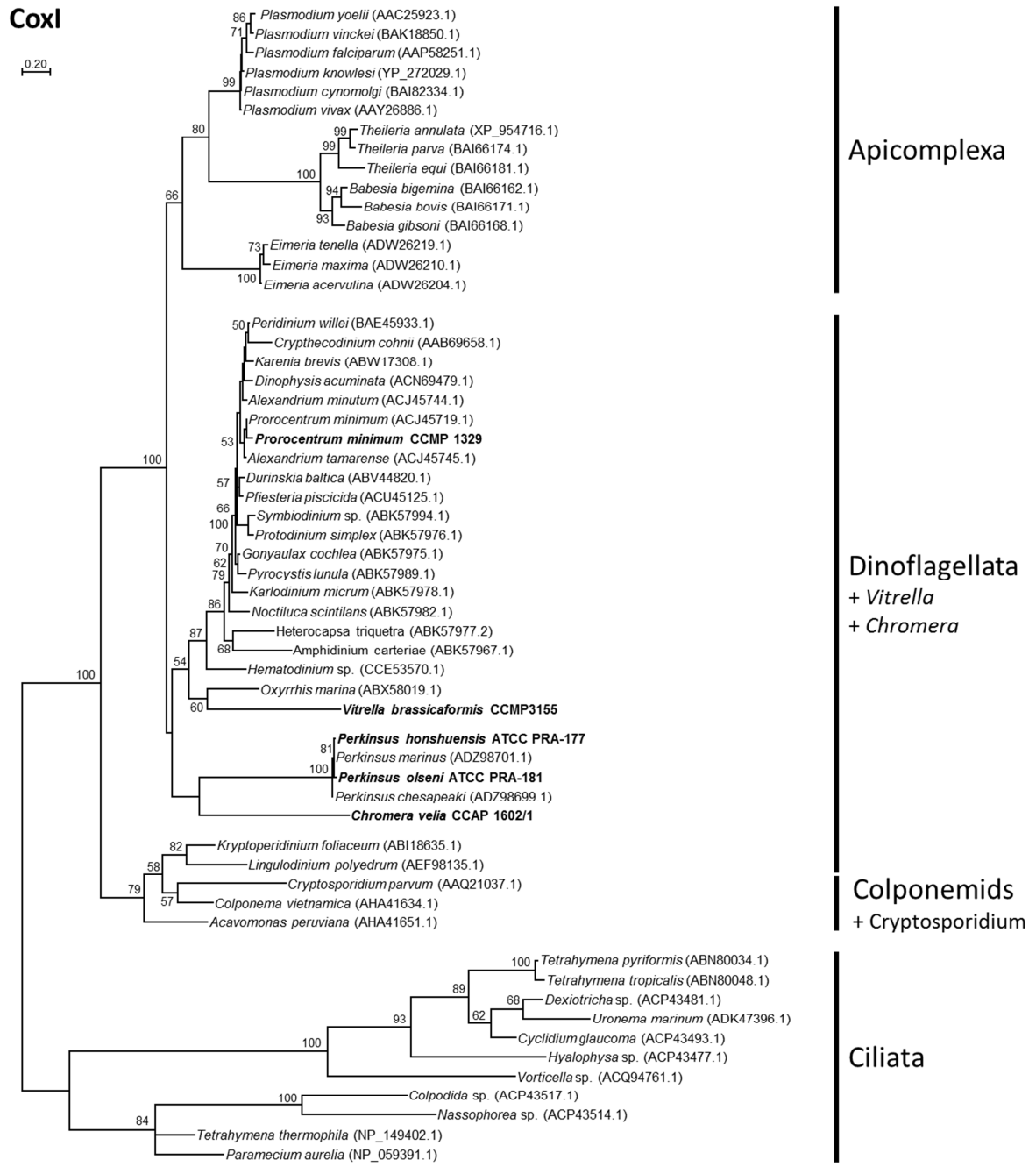


Figure 28: Phylogenetic maximum likelihood RAxML tree (WAG+F+Γ4) based on 57 alveolate *coxI* protein sequences with an alignment length of 789 aa. Sequences from newly established genomes of this thesis are marked in bold. The ciliates were chosen as outgroup.

Mitochondrial Pathway Inventory in Early-branching Alveolates

The discovery of a mitochondrial genome harboring only one protein-coding gene raises the question if *C. velia* cells still incorporate fully functional mitochondria. Concerning this issue a metabolic inventory of mitochondrial pathways was established based on a marker composition for alveolates published by Danne et al. (2013). Furthermore, the marker information of the other early-branching alveolates with strongly reduced mitochondrial genomes, *V. brassicaformis* and *P. olseni*, was added. The analysis of *P. minimum* served as an internal reference (Table S7). The transcriptomic as well as genomic data were combined in order to identify the presence of introns and thus a nuclear localization of markers. The analyses revealed full functional mitochondria in *P. olseni*, *V. brassicaformis* as well as in *C. velia*, which have successfully transferred the bulk of its mitochondrial genomic content to the nucleus of the cell and established an efficient protein import system to maintain the organelle and its functions (Table S7). The further description is focused on the respiratory chain and the Fe-S cluster assembly apparatus reflecting the pivotal core functions of mitochondria (Chapter 1.5.1).

(1) Oxidative Phosphorylation in Chromerids and *P. olseni*

The absence of the central protein NADH dehydrogenase of the complex I is a shared trait between all myzozoan species. Due to the fact that it is also absent in early-branching species confirms the hypothesis that this enzyme was already absent in a common ancestor of the Myzozoa. Instead complex I, the alternative NADH dehydrogenase (NDH2) homologous to fungi and land plants adopts the function of reducing coenzyme Q (Chapter 1.5.1, Table S7). Additionally, complex II which is not involved in the formation of a proton gradient supports the reduction of ubiquinone.

Whereas, *P. olseni* and *V. brassicaformis* inhabit all markers for the complex III, surprisingly *C. velia* seems to have abandoned this system. That does not result in a dead-end for the electron transport chain because *C. velia* encodes for an alternative oxidase (AOX), which uses the electron transport chain to reduce directly oxygen to water as a bypass of complex III and complex IV. The presence of AOX was not integrated into the previous dataset of Danne et al. (2013). Thus, I performed blastp and blastn analyses searching for the AOX in the other alveolate species. The AOX protein seems to have been already present in a common ancestor of alveolates because of its omnipresence in all considered organisms (Table S7). However, an AOX sequence could not be identified via any blast search in apicomplexan parasites with the exception of *Cryptosporidium parvum* and *Gregarina niphandrodes*, although the

presence of this protein was hypothesized in all apicomplexan parasites by Roberts et al. (2004) due to effects on inhibitor experiments of the AOX. The alternative pathway via the AOX was shown to be induced by exogenous stress factors like e.g. oxidative stress and it somehow decreases the formation of reactive oxygen species, which are toxic for the cell (Maxwell et al. 1999). This would explain its presence in photosynthetic and aerobic alveolates. In parasites, the AOX protein expression could be enhanced during the invasion in their hosts, because there they are exposed to multiple stress factors induced by the immune reaction of the host.

In all early-branching alveolates studied in this work one protein (COX3) of the otherwise present complex IV is missing. This discovery conforms to the apparent absence of this protein in ciliates. Perhaps, it could be replaced by a non-homologous nuclear-encoded protein with similar functions in these lineages or its sequence is too divergent to be found by blast analyses (Flegontov et al. 2015).

The presence of complex IV and the complete set of proteins for a functional ATPase complex in all early-branching myzozoan species (Table S7) indicates that the oxidative phosphorylation pathway of *C. velia* is potentially divided into two functional parts: (1) one pathway which is not involved in the formation of a proton gradient including the central alternative enzymes NDH2 and AOX; and (2) a second pathway based on the formation of a proton gradient by complex IV and on the production of ATP induced by this gradient with the ATPase complex. My hypotheses are supported by the contemporaneous mitochondrial studies of Flegontov et al. (2015), who accessorily linked the presence of the additional dehydrogenase Cytochrome b2 (COB2; Table S7) and another enzyme (L-galactono-1,4-lactone dehydrogenase) as electron donors to the omnipresent Cytochrome c (Table S7), which is oxidized by complex IV (Chapter 1.5.1).

(2) Fe-S cluster assembly by the Isc system

Most of the marker proteins of the Isc system for iron-sulfur (Fe-S) cluster assembly are present in all considered alveolate species indicating also the functionality of this pathway in mitochondria of early-branching myzozoan lineages (Table S7). Fe-S clusters are *inter alia* important cofactors within the enzymes of the electron transport chain (ETC) in the oxidative phosphorylation pathway. As already shown in Chapter 1.5.1, Fe-S clusters are integrated into complex I, complex II and complex III of the ETC, thus in *C. velia*, the Fe-S clusters are only needed for the complex II.

3.3 Peroxisomal Evolution in Alveolata

"*It became obvious that peroxisomes are highly dynamic organelles that rapidly assemble, multiply and degrade in response to metabolic needs.*"

(Schrader, Fahimi 2008)

The presence of peroxisomes in the superensemble Alveolata was extensively studied for the ciliates, since its early discovery in 1965 by Baudhuin (e.g. Hogg 1969; Müller 1973; Blum 1982). In contrast, the presence of this organelle was controversially discussed in Apicomplexa (Ding et al. 2000; Kaasch, Joiner 2000). However, it was a long time assumed that a peroxisome is absent in this phylum (Schlüter et al. 2006; Gabaldon 2010). In dinoflagellates transmission electron micrographs indicated the presence of a peroxisome organelle (Bibby, Dodge 1973) and recent transcriptomic analyses identified two metabolic genes of a peroxisome-specific pathway (Butterfield et al. 2013). In this chapter of my PhD thesis I characterized the peroxisome organelle in four key species of the myzozoa (*C. velia*, *V. brassicaformis*, *P. olseni* and *P. minimum*) in comparison to the ciliate reference organism *T. thermophila* based on diagnostic proteins. Apart from the reconstruction of peroxisome evolution in alveolates, I wanted to carve out the presence or absence of this organelle in the apicomplexan parasites. For this reason, *in silico* analyses of peroxisomal markers were conducted in specific transcriptomic and genomic databases for *C. parvum*, *T. gondii* and *P. falciparum* (CryptoDB 27, [Heiges et al. 2006]; ToxoDB v12.0, [Gajria et al. 2008]; PlasmoDB v12.0, [Aurrecoechea et al. 2009]) as well as NCBI-located datasets of other whole-genome sequenced, parasitic Apicomplexa.

3.3.1 The Presence of Peroxins in Myzozoan Key Species Indicates the Presence of Peroxisomes in Dinoflagellata and Apicomplexa

In the established alveolate transcriptomes (*C. velia*, *V. brassicaformis*, *P. olseni* and *P. minimum*) and initially in databases of other alveolate key species (*C. parvum*, *T. gondii* and *P. falciparum*) it was searched for peroxisome-specific proteins, the peroxins, in order to describe a putative peroxisome organelle *in silico*. Tblastn searches were conducted with appropriate eukaryotic query sequences (e.g. *A. thaliana*, *Caenorhabditis elegans*, *Dictyostelium discoideum*, *Homo sapiens* or *S. cerevisiae*), but mainly with the ciliate reference *T. thermophila*. 14 out of 16 potential peroxins could be detected within all the myzozoan key species with an alveolate core set of six peroxins (Pex1, Pex4, Pex5, Pex6, Pex7 and Pex11; Figure 29). The number of peroxins in the ciliate reference *T. thermophila* amounts to eleven enzymes, whereas in the dinoflagellate species 13 peroxins in *P. olseni* and

ten peroxins in *P. minimum* were detected (Figure 29). Both chromerid transcriptomes, *C. velia* and *V. brassicaformis*, comprise 13 to 14 peroxins, and their close apicomplexan relative *T. gondii* incorporates eleven peroxins, respectively (Figure 29). This analysis is in striking contrast to previous claims that apicomplexan species are lacking peroxisomes (Schlüter et al. 2006; Gabaldon 2010). However, in the other two parasitic apicomplexan species (*C. parvum* and *P. falciparum*) only homologous genes to *pex4* and *pex22* sequences could be identified inferring that in these species a peroxisome was lost secondarily. That would explain the absence of all other peroxins in these two species. The peroxins Pex13, Pex15 and Pex26 are in general absent in all alveolate key species. This analysis represents the first evidence of the presence of peroxisomes in all alveolate phyla.

Importomer (Pex5, Pex7, Pex14)

The receptors Pex5 and Pex7, are important for the import of peroxisomal matrix proteins with peroxisomal targeting sequences, the C-terminal PTS1 and the N-terminal PTS2 (Williams and Distel 2006, Chapter 1.5.3). They were detected in six analyzed alveolate datasets, but they were absent in the inferred aperoxisomal species *C. parvum* and *P. falciparum*. The potential transmembrane docking partner Pex14 was identified in the chromerids (*C. velia* and *V. brassicaformis*) as well as in the apicomplexan parasite *T. gondii* and in the dinoflagellate parasite *P. olseni*. Another potential docking partner (Pex13) was universally absent in alveolates. Due to the fact that any previous described docking partner for Pex5 and Pex7 is missing in the ciliate reference *T. thermophila*, ciliates must have developed another non-homologous docking station for the two PTS receptors (Figure 29).

Exportomer (Pex1, Pex6, Pex2, Pex10, Pex12)

The ATP-dependent exportomer has a central function in recycling and degradation of the cargo receptor Pex5 based on its ubiquitination level (Smith, Aitchison 2013). The ATP suppliers of the exportomer, Pex1 and Pex6, were again detected in all six alveolate species, the supposed aperoxisomal species *C. parvum* and *P. falciparum* excluded. In contrast, a transmembrane anchor for the Pex1-Pex6-complex, either a homologous sequence to the human Pex26 or the fungal Pex15 protein, was universally absent. RING finger domain proteins are involved in the ubiquitination cascade of the Pex5-exportomer (Chapter 1.5.3). Thus, for the degradation process Pex2 was found in all alveolate species, the parasitic Apicomplexa excluded. For the monoubiquitination as a degrading marker of Pex5, Pex2 is supported by the ubiquitin-conjugating enzyme Pex4 and its membrane anchor Pex22. Both, Pex4 and Pex22, were identified in all alveolate datasets even in aperoxisomal Apicomplexa,

whereas Pex22 is missing in the ciliate *T. thermophila*. Likewise involved in the degradation of Pex5 is Pex10, which is only present in *T. thermophila* and the photoautotrophic Apicomplexa (*C. velia* and *V. brassicaformis*) in this study. Pex12 is implicated into the recycling of Pex5 and could be detected in all alveolate key organisms, except *P. minimum*, *C. parvum* and *P. falciparum* (Figure 29).

Membrane Assembly and Peroxisome Organization (Pex3, Pex16, Pex19, Pex11)

With regard to the assembly of the peroxisomal membrane, a second protein import system for peroxisomal membrane proteins (PMPs) is required. In the focus of this import system is the cytosolic receptor protein Pex19, which recognizes the membrane PTS of the PMPs and acts like a chaperone (Jones et al. 2004). Together with the ciliate reference *T. thermophila* this receptor could be identified in the transcriptome of *V. brassicaformis* and in the dinoflagellate key species, whereas the docking partner of Pex19, Pex3, was detected in the myzozoan species (*C. parvum* and *P. falciparum* excluded), but was absent in this ciliate (Figure 29). The absence of Pex19 in the chromerid *C. velia* was confirmed by a cross-check with another whole-genome dataset of Woo et al. (2015) using the Pex19-homolog of *V. brassicaformis* as a query sequence. Thus, Pex19 must have been replaced by another functional protein in all other peroxisome-harboring Apicomplexa to ensure the import of integral PMPs into the membrane of this organelle.

Concerning the absence of the docking partner Pex3 in some species, homologs to the Pex19 docking partner in mammals, Pex16, were detected in *T. thermophila* as well as in *V. brassicaformis*, *T. gondii* and *P. olseni* (Figure 29). One integral membrane protein, which is imported by the receptor protein Pex19, is Pex11 that has a crucial function in peroxisomal fission (Fagarasanu et al. 2007). Based on its importance in peroxisome organization this gene was detected in all alveolate datasets, except for *C. parvum* and *P. falciparum.*

Transport through the Peroxisomal Membrane

Integral peroxisomal membrane proteins (PMPs) can serve as transporters for metabolites that cross the peroxisomal membranes. The interchange of metabolites between the cytosol and the peroxisomal matrix is only poorly studied, but three PMPs could be identified and are highlighted in this study (Figure 29). The pore-forming PMP22, whose pore is likely permeable for small solutes with a size up to 300 Da, is also involved in lipid metabolism (Antonenkov, Hiltunen 2012). This protein could be detected in the chromerids, *P. olseni* and the ciliate reference *T. thermophila*. Another carrier for metabolites is the importer of ATP,

PMP34, which is involved in the β-oxidation of fatty acids. This protein could be identified in all peroxisome-harboring alveolate key species of this study. The same is true for the largest 70-kDa carrier protein PMP70, which is a typical ABC transporter for the import of fatty acids into the peroxisome.

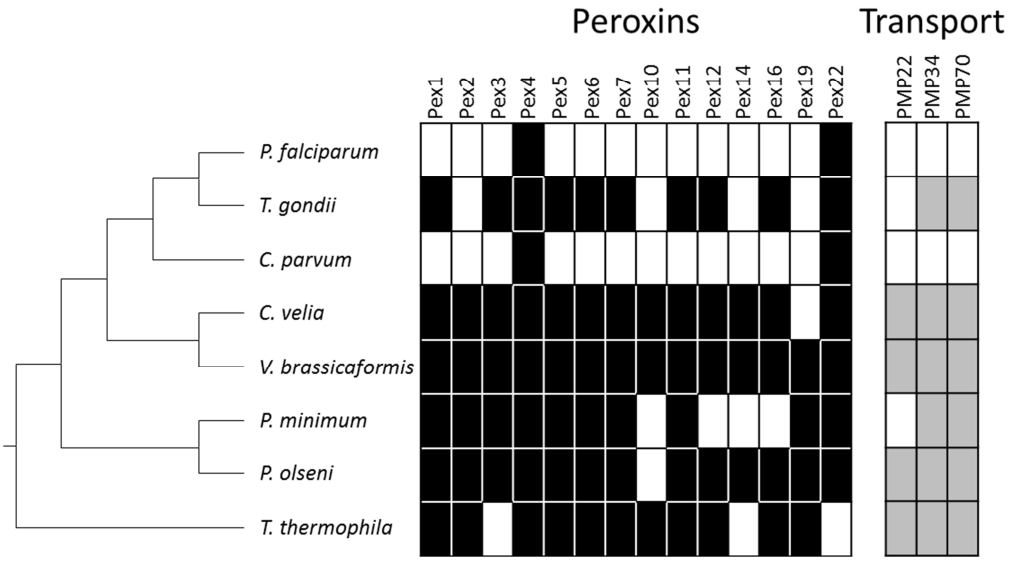


Figure 29: Presence/absence of the peroxisome-specific marker proteins, the peroxins (PEX), as well as peroxisomal membrane proteins (PMPs), which contribute to the membrane permeability of the organelle. The presence of the respective markers are marked by a colored (black = peroxins; grey = PMPs), absence by a colorless box.

Phylogenetic Analyses of PEX1 and PEX5

Peroxin phylogenies were constructed in order to deduce the evolution of peroxisomes in protists. Unfortunately, peroxins are not very suitable for phylogenetic analyses, because of their short length, which come along with divergent alignments and not enough amino acid positions for a reliable phylogenetic tree. Two out of the six core set peroxins, Pex1 and Pex5, were selected which provide the most informational content (Figure 30). Since all peroxisomal proteins are encoded in the nucleus, one would expect phylogenies following the pattern of the species tree and thus the monophyly of the Alveolata as the null hypothesis.

The phylogenetic RAxML tree of **Pex1**, a homolog of the ERAD/SELMA system protein CDC48, shows the expected monophyly of alveolates, which is only weakly supported (Figure S15). The low support in favor of their monophyly may be a combined effect of their higher evolutionary rate in comparison to other groups in this tree and the low number of positions. For example, the comparable fast monophyletic group of kinetoplastids (100% BS) is attracted via LBA and forms a sister group to the myzozoan sequences (Figure 30/Figure S15). The accelerated evolutionary rate of alveolates results in a reduction of the alignment

size to only 179 comparable positions. The photosynthetic Apicomplexa, *C. velia* (KR704675) and *V. brassicaformis* (KR704726), show a serial branching pattern (45% BS and 40% BS) as the sister to the monophyletic groups of parasitic Apicomplexa and Dinoflagellata (100% BS and 97% BS). Comparable to the mitochondrial *coxI* phylogeny in chapter 3.3.2 the dinoflagellates are paraphyletic and the *Perkinsus* species have a low affinity to an incomplete sequence of the fast evolving haptophyte *E. huxleyi* (Figure 30/Figure S15). This phenomenon could be likewise induced by a long branch attraction artefact (LBA). However, the tree shows a eukaryotic origin of this gene (unikonts as outgroup) and a common ancestry in alveolates.

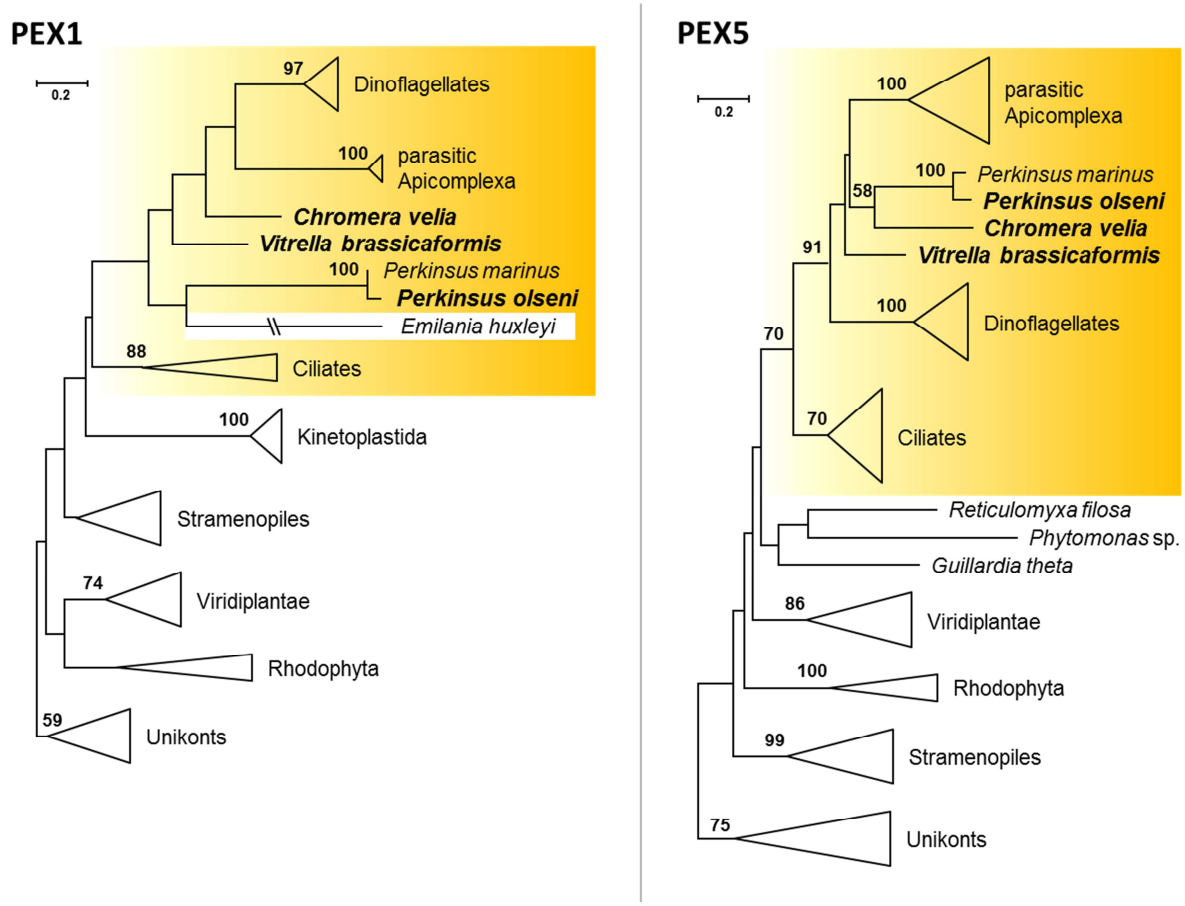


Figure 30: Phylogenetic RAxML analysis (LG+F+Γ4) for the two peroxins Pex1 and Pex5. The Pex1 tree is based on 60 sequences and 179 aa positions, and the Pex5 tree on 59 sequences with 205 aa positions. Well-defined taxonomic units are collapsed to triangles. The myzozoan sequences of my newly established transcriptomes are highlighted in bold and a larger font size. Positions of members of the superensemble Alveolata are highlighted with a yellow box. The complete phylogenetic trees are shown in Figure S15 and Figure S16.

The phylogenetic analysis of the second peroxisome-specific marker **Pex5** exhibits a higher resolution and also more moderate evolutionary rates. In this case the monophyly of alveolates is stronger supported, which is in agreement with our null hypothesis (Figure 30/Figure S16; 70% BS). The alveolate subtrees of ciliates, autotrophic dinoflagellates and

parasitic Apicomplexa are monophyletic (70% BS; 100% BS; 100% BS; Figure 30/Figure S16) and even the group of Myzozoa was highly supported (91% BS; Figure 30/Figure S16). In order to figure out the precise position of the photoautotrophic Apicomplexa, the alveolate group was reanalyzed separately (Figure S17). In the original tree (Figure S16) *V. brassicaformis* was the sister to a group consisting of *C. velia*, *Perkinsus* spp. and the parasitic Apicomplexa. This deep position is only weakly supported (32% BS), whereas the support of the association of *Chromera* with the *Perkinsus* species is higher (58% BS). The subanalysis shows the same result with higher bootstrap support for the *Chromera-Perkinsus*-clade (84% BS, Figure S17), but a weaker position of *V. brassicaformis* at the basis to the Apicomplexa-*Chromera-Perkinsus*-clade (30% BS). This result could either represent a phylogenetic artifact or it is a valid horizontal transfer of this peroxin from a chromerid to the genus *Perkinsus*. Comparable to Pex1, Pex5 is of eukaryotic origin (unikonts as outgroup) and supports a common ancestry in alveolates.

3.3.2 Presence of Crucial Peroxisomal Metabolic Markers in Alveolate Key Species

As peroxisomes are implicated in numerous metabolic pathways, I highlighted only the pathways in this subchapter, which account for the presence of this compartment in the cell. Therefore, I focused on the glyoxylate cycle, the β-oxidation of fatty acids, the ether phospholipid biosynthesis and the antioxidant system filtering out the most important markers (Figure S1; Chapter 1.5.3).

Glyoxylate Cycle (GC)

Based on the discovery of Butterfield et al. (2013) that dinoflagellates harbor the two key enzymes ICL and MLS of the peroxisome-specific glyoxylate cycle, the search for the genes of these proteins via tblastn in all newly established transcriptomes of Dinoflagellates (*P. olseni* and *P. minimum*) and even of the Apicomplexa (*C. velia* and *V. brassicaformis*) was conducted.

According to de Duve and Baudhuin (1966) the presence of all five enzymes of the glyoxylate cycle could be confirmed in the reference ciliate *T. thermophila*, which served as a source of query sequences for the local blast analyses in the transcriptomes. Transcripts of the ICL and the MLS were detected in all four alveolate transcriptomes, but not in the parasitic apicomplexan species *T. gondii* and *P. falciparum* (Table 6). The presence of these two key enzymes indicates the existence of the glyoxylate cycle in these four species. The three remaining glyoxysomal proteins CS, ACO and MDH were found in all alveolate species,

whereas in the parasitic Apicomplexa isoenzymes of these enzymes could be identified in the mitochondrial TCA cycle by demonstrating their mitochondrial targeting sequences with the prediction program TargetP (Emanuelsson et al. 2000). Moreover, prediction programs for peroxisomal targeting sequences (PTS1 and PTS2), iPSORT (Bannai et al. 2002) or the prediction tool of PeroxisomeDB (Schlüter et al. 2010), allowed to detect the N-terminal and C-terminal PTS sequences in MLS and CS of *P. olseni*, in ICL and MLS of *C. velia* and in the CS of *V. brassicaformis*. Surprisingly, a C-terminal PTS sequence was found in the MDH enzyme of *P. minimum*, which could indicate a different spatial distribution of the glyoxylate enzymatic reactions than in land plants and fungi (Table 6).

Table 6: Presence/absence of the glyoxysomal marker proteins. Presence is indicated by a check mark, absence by a dash. Peroxisomal targeting sequences (PTS) are marked by a superscripted 1 (C-terminal PTS1) and 2 (N-terminal PTS2). Other superscripted characters: M, mitochondrial localization; C, cytosolic localization. Abbreviations for organisms: Tt, *Tetrahymena thermophila*; Po, *Perkinsus olseni;* Pm, *Prorocentrum minimum*; Vb, *Vitrella brassicaformis*; Cv, *Chromera velia*; Cp, *Cryptosporidium parvum*; Tg, *Toxoplasma gondii*; Pf, *Plasmodium falciparum.*

Protein	Abbr.	Tt	Po	Pm	Vb	Cv	Cp	Tg	Pf
Isocitrate lyase	ICL	✓	✓	✓	✓	✓^{1}	–	–	–
Malate synthase	MLS	✓	✓^{1}	✓	✓	✓^{1}	–	–	–
Citrate synthase	CS	✓	✓^{2}	✓	✓^{2}	✓	–	✓^{M}	✓^{M}
Aconitase	ACO	✓^{C}	✓^{C}	✓^{C}	✓^{C}	✓^{C}	–	✓^{C}	✓^{C}
Malate dehydrogenase	MDH	✓	✓	✓^{1}	✓	✓	✓	✓^{M}	✓^{M}

Antioxidant System (AoxS)

The analysis of antioxidant enzymes in alveolate key species revealed that the previous peroxisomal key marker catalase (Chapter 1.5.3; CAT) is not mandatory for the presence of the peroxisome given that other oxidoreductases are adopting or replacing its function. A catalase in *T. thermophila* (PeroxisomeDB, (Schlüter et al. 2006), *V. brassicaformis* and *T. gondii* could be detected (Figure 31), which was also already documented for *Toxoplasma* in the literature by Ding et al. (2000). In *C. velia* and the dinoflagellates (*P. olseni* and *P. minimum*) the catalase was potentially replaced by the peroxiredoxins PRDX1 and PRDX5 (EC 1.11.1.15). The presence of the glutathione S-transferase kappa 1 (GSTK1; EC 2.5.1.18) is restricted to chromerids and dinoflagellates and the superoxide dismutase (SOD; EC

1.15.1.1) was detected in *T. thermophila*, *P. olseni* and *T. gondii* which are both also responding to oxidative stress. Surprisingly both, the SOD and PRDX1, were identified in the aperoxisomal species *C. parvum* and *P. falciparum* (Figure 31). A possible explanation for the presence of these enzymes in parasites would be the protection of their cells against oxidative damage by defense mechanisms of the host cells (Clark, Hunt 1983). The presence of the H_2O_2-producing enzymes of the fatty acid oxidation pathway ("FAO" Figure 31; ACOX), amino acid metabolism ("AAP" Figure 31; PAOX, PIPOX) and purine metabolism ("P" Figure 31; XDH) in most of the peroxisomal alveolates indicates the internal presence of antioxidant enzymes.

.

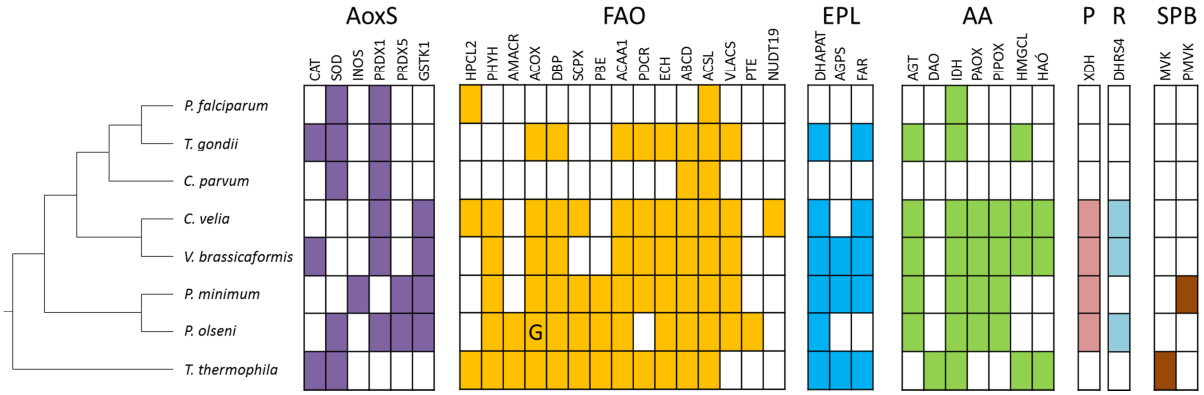


Figure 31: Presence/absence of key markers for peroxisomal metabolic pathways inspired by the compilation of a peroxisomal pathway map in KEGG (Kanehisa, Goto 2000) in eight alveolate key organisms. The presence of a respective marker is pigmented in the color which corresponds to the illustration in Figure S1. The ACOX sequence found in the genome of *P. olseni*, but was absent in the transcriptomic dataset, is marked by a 'G'. Abbreviations: AoxS = antioxidant system; FAO = fatty acid β-oxidation; EPL = etherphospholipid biosynthesis; AA = amino acid metabolism; P = purine metabolism; R = retinol metabolism; SPB = sterol precursor metabolism.

Fatty acid β-oxidation (FAO)

The catabolism of fatty acids is represented by three enzymes, which are found in all peroxisomal alveolates, but which are absent in the aperoxisomal key species *C. parvum* and *P. falciparum* (DBP, ACAA1, ECH; Figure 31). The H_2O_2 - enzyme ACOX (EC 1.3.3.6) is found in all peroxisomal alveolate key organisms, but not in *P. olseni* (Figure 31). However, the ACOX could be identified in the close relative *P. marinus* (XP_002774062), indicating that the absence of ACOX is just a false negative in the transcriptomic dataset of *P. olseni*. Actually, a local tblastn analysis with the *P. marinus*-query in the genomic dataset of *P. olseni* revealed one homologous sequence ('G', Figure 31). Homologs of three enzymes (HPCL2, ABCD and ACSL) are also found in aperoxisomal apicomplexan species, which might be relicts of a former existing peroxisome (Figure 31). Analyses with targeting prediction programs revealed a possible mitochondrial localization of these enzymes (data not shown),

which is in agreement with a supposed flexibility in the localization of the fatty acid catabolic pathway (Shen, Burger 2009).

Etherphospholipid biosynthesis (EPL)

Corresponding to the detected fusion protein "TtFARAT" in the ciliate *T. thermophila* (Chapter 1.5.3; Dittrich-Domergue et al. 2014), a homologous fusion-protein between the fatty acyl-CoA reductase (FAR; EC 1.2.1.84) and the dihydroxyacetone phosphate acyltransferase (DHAPAT; EC 2.3.1.42) could be identified in *P. minimum* and *C. velia*. Whereas the DHAPAT could be shown in all peroxisomal alveolates, the FAR seems to be absent in the transcriptomic dataset of *P. olseni* and was also not detectable in the genomic dataset. Thereby, the DHAPAT sequence is not well conserved, but it is still detectable. The enzyme alkyldihydroxyacetone phosphate synthase (AGPS; EC 2.5.1.26) seems to be only present in *T. thermophila*, *P. minimum* and *V. brassicaformis* (Figure 31). In comparison to the highly divergent sequence of the DHAPAT within the fusion protein, it is possible that the AGPS sequence is so much derived that it cannot be identified by blast analyses.

Phylogenetic Analyses of Metabolic Markers

I selected suitable diagnostic markers to reconstruct the metabolic evolutionary history in peroxisomes. The glyoxysomal proteins ICL and MLS are suitable markers because of their peroxisome-specific metabolic function and their conservation throughout ciliates, dinoflagellates and photoautotrophic Apicomplexa. Numerous test runs with other metabolic markers ended up in phylogenetic trees with insufficient resolution. Anyway, the ACAA1 of the fatty acid oxidation pathway asserted as a suitable marker.

Since all peroxisomal enzymes are nuclear-encoded, I would again expect a topology highly similar to that of the species tree (Figure 13) and thus a monophyly of the alveolate superensemble. In contrast the phylogenetic analyses often revealed a paraphyly of the alveolates with different origins (Figure 32; Figure 33) rejecting the hypothesis of a common ancestry of the glyoxylate cycle in alveolates.

Based on the original phylogenetic tree of **ICL** (Figure S18), monophyletic taxonomic units (Fungi, Streptophyta, Ciliates, *Perkinsus* sp., Oomycetes, Dinoflagellates) and a subtree (Stramenopiles with *Oxyrrhis* and *Emiliania*; 54% BS) were collapsed to triangles to form a schematic tree shown in Figure 32. Streptophyte sequences are the basis of this tree and support a eukaryotic origin of this gene. The ciliates are the sister group to a big subtree consisting of CASH group members. Within this subtree the dinoflagellates are paraphyletic:

the majority of dinoflagellates with our study organism *P. minimum* are monophyletic as a sister group of the oomycetes (99% BS), whereas another solidly supported subtree (98% BS) incorporates duplicated sequences of the genus *Perkinsus* including our key species *P. olseni* (Figure S18). The Perkinsea have a weak affinity to the ICL sequence of the cryptophyte *G. theta* (45% BS). However, the dinoflagellate sequence of *O. marina* branch together with the photosynthetic stramenopiles (54% BS). The photosynthetic apicomplexan algae, *C. velia* and *V. brassicaformis*, cluster together (63% BS) and form the sister group to the weak supported subtree of Stramenopiles, *O. marina*, *E. huxleyi*, the Perkinsea and *G. theta* (Figure 32; Figure S18).

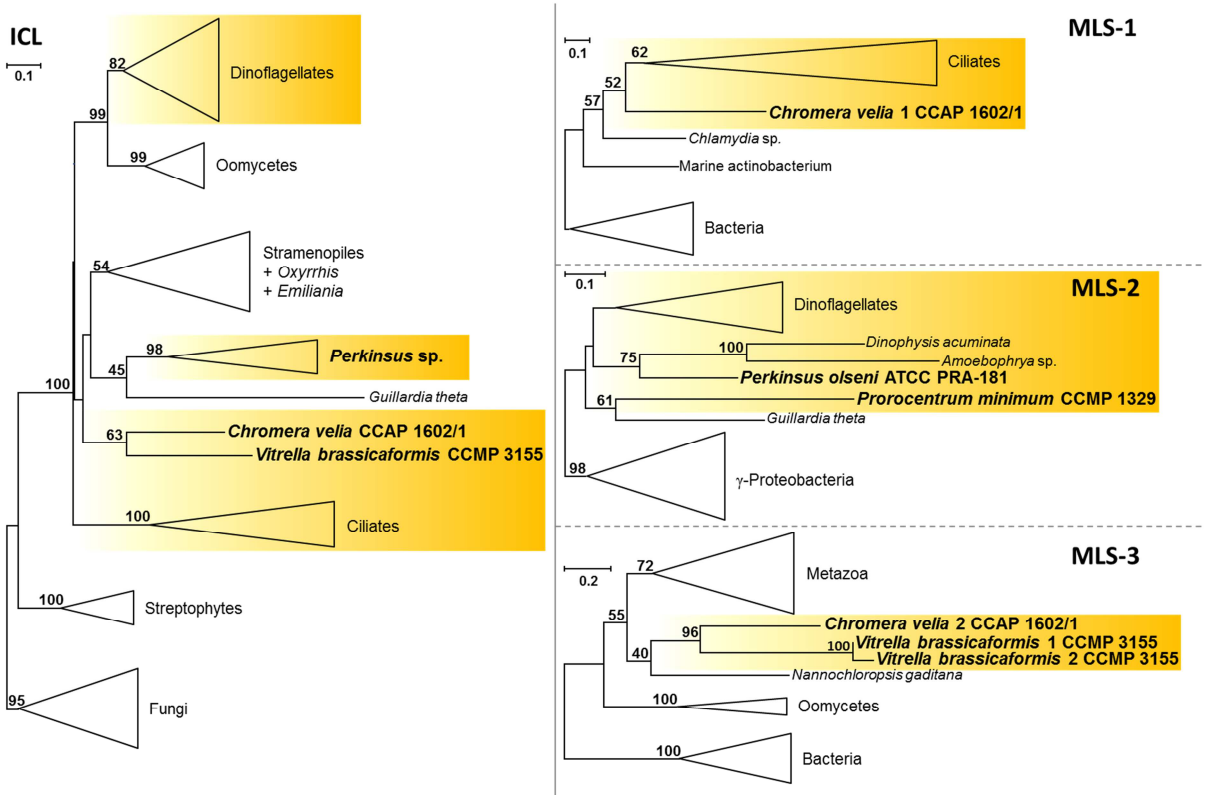


Figure 32: Phylogenetic RAxML analyses based on a LG+F+Γ4 model for the two glyoxysomal markers ICL and MLS. The ICL tree was calculated based on 47 sequences with 398 aa positions. The MLS tree is splitted into three subanalyses due to different bacterial origins. MLS-1 incorporates 34 sequences with 415 aa positions, MLS-2 20 sequences and 513 aa positions, and MLS-3 45 sequences and 319 aa positions. Clearly defined groups are collapsed to triangles. My myzozoan key species are highlighted in bold and with a larger font size. The position of the alveolate species is accentuated by a yellow box. The complete phylogenetic trees are shown in Figure S18 - Figure S21.

The phylogenetic analyses for the second glyoxysomal marker **MLS** document at least three independent recruitments from bacteria. The alveolate species are found in different regions of a large original tree (data not shown), because the gene has different bacterial origins. For that reason three subtree analyses were conducted (Figure 32; Figure S18 - Figure S21). Once again monophyletic taxonomic units (bacteria, ciliates, dinoflagellates, Metazoa and

oomycetes) were collapsed to triangles demonstrated in Figure 32 by the three schematic trees. The ciliates are included in the first subtree together with a paralogous sequence of *C. velia* (52% BS, MLS-1 Figure S19, Figure 32). A chlamydial sequence is closely associated to this "alveolate cluster" (57% BS) followed by a broad diversity of other bacterial sequences forming the outgroup of this subtree. The MLS gene of the dinoflagellates is of γ-proteobacterial origin and is found in the second subtree (MLS-2 Figure S20, Figure 32). Thereby the *P. minimum* sequence is affiliated to the sequence of the cryptophyte *G. theta* (61% BS). Finally, in the third subtree the sequences of the photoautotrophic apicomplexan species, *C. velia* and *V. brassicaformis*, form a monophyletic group (96% BS) and have the stramenopile *Nannochloropsis gaditana* as a sister (40% BS). This subgroup clusters together with the Metazoa (55% BS) and the oomycetes (100% BS, MLS-3 Figure S21, Figure 32). The bacterial outgroup comprise δ-proteobacteria, acidobacteria and actinobacteria.

In combination with the results of the peroxin phylogenies (Chapter 3.4.1, Figure 30) incongruences could be detected in the evolutionary inheritance of the peroxisome organelle. It seems to be that the presence of the organelle itself is highly conserved and that it was already present in a common ancestor of all alveolate species, but that genes with peroxisomal metabolic function were replaced later depending on the needs of the cell. Thus horizontal gene transfer by bacterial sequences or other sources are very likely exemplified in the phylogenies of the glyoxysomal genes (Figure 32).

However, the phylogenetic reconstruction of the **ACAA1** marker of the fatty acid degradation pathway is more consistent with our null hypothesis. In this case also parasitic Apicomplexa could be included into the analysis. The monophyly of all three alveolate phyla is well supported (ciliates 61% BS; dinoflagellates 67% BS; Apicomplexa 88% BS; Figure 33; Figure S22) and the alveolate sequences are weakly associated with fungi. Thereby, the group of fungi and ciliates forms the sister to the monophyletic Myzozoa (82% BS; Figure 33). However, the grouping of ciliates and fungi is not resolved and a subanalysis of a reduced dataset incorporating only alveolates, stramenopiles and fungi shows the expected sister branching of ciliates to the Myzozoa, although two fast evolving ciliate sequences are again spuriously attracted to fungal sequences (Figure S23). Perhaps due to LBA artifacts, the parasitic group of *Eimeria* sp. is mispositioned in this tree (Figure 33). The reanalysis of the ACAA1 marker revealed again a serial branching order of the chromerids with *V. brassicaformis* as the deepest apicomplexan branch and a position of *Eimeria* corresponding to the species tree (Figure S23; Figure S1). The presence of the fatty acid oxidation pathway

in parasitic Apicomplexa and its considerable conservation indicates that this pathway has an important status in the maintenance of the peroxisome organelle.

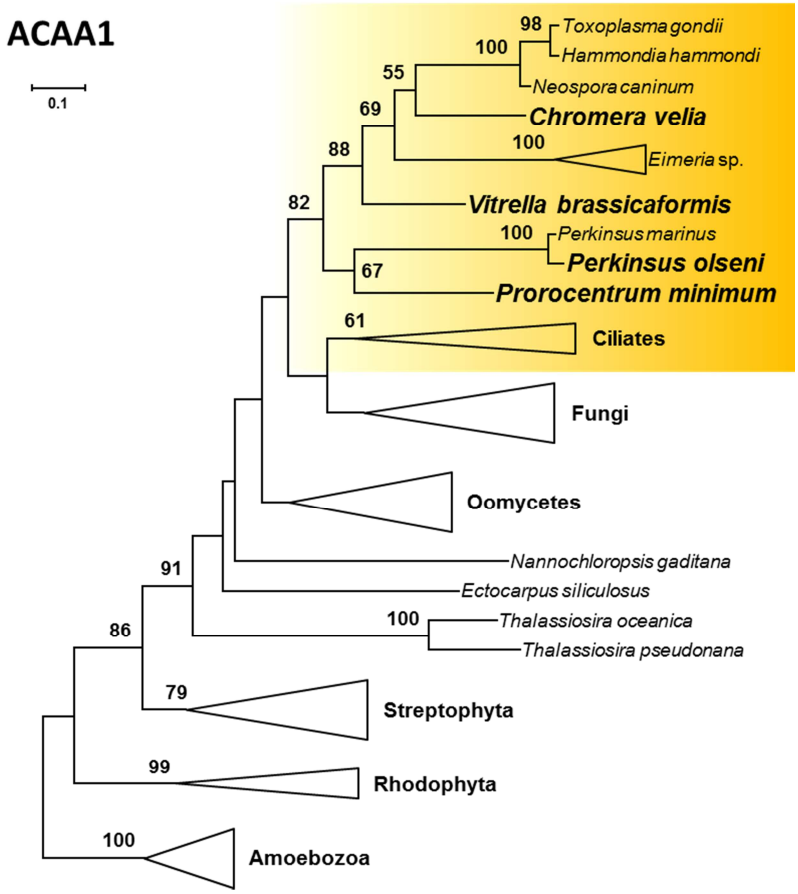


Figure 33: Phylogenetic maximum likelihood RAxML tree (LG+F+Γ4) based on 57 ACAA-1 sequences and 326 aa positions. Clearly defined groups are collapsed to triangles. The newly established myzozoan data are highlighted in bold and with a larger font size. The position of the alveolate species is accentuated by a yellow box. The complete phylogenetic tree is shown in Figure S22.

3.3.3 Comparative Peroxisomal Inventory in Alveolates

The inventory of peroxisomal markers in the current PhD thesis represents the first clear-cut evidence for the presence of a peroxisome in Apicomplexa and dinoflagellates. The peroxisomal localization was confirmed by the detection of peroxisomal targeting sequences (PTS) in numerous metabolic key genes (Table S8) with a PTS1 consensus sequence of [S/A/G]-[R/K/N]-[L/M/A] in alveolates whereby the C-terminal ending –SRL was the most common PTS1 signal. The identification of the N-terminal PTS2 sequence is challenging due to its low conservation level. However, its presence is indicated in some metabolic enzymes by appropriate prediction programs (PeroxisomeDB; Schlüter et al. 2010; PSORTII, Nakai, Horton 1999), but these results must be considered with care.

Consistent with plastids and mitochondria, the peroxisome seems to experience a kind of gradual reductive evolution in terms of molecular functions this organelle processes which is dependent on the environmental adaptation level of the respective organism. Thus, comparisons of the peroxisomal nucleus-encoded gene content of the photoautotrophic alga *C. velia* and its close parasitic apicomplexan relative *T. gondii* revealed 24 metabolic genes to 16 genes, respectively. In general, parasitic organisms are dependent on the interaction with their hosts, thus multiple metabolic functions and reaction intermediates can be replaced by this interaction. For this reason, the malaria agent *P. falciparum* obviously completely lost its peroxisome (only five putative metabolic enzymes found, Figure 31)**.**

Although, *C. velia* shows the highest number of detected metabolic markers, the closest relative *V. brassicaformis* displays the largest set of peroxisomal markers, which could be identified in alveolates so far. Thus, this organism would provide the perfect reference organism to study the evolution as well as the function of peroxisomes in alveolates. All acquired information of the *V. brassicaformis* peroxisome are summarized in a pathway map inspired by the KEGG PATHWAY database (Kanehisa, Goto 2000, Figure 34).

Investigations in the alveolate phylum Apicomplexa revealed that the peroxisome organelle must have been lost at least twice in parasitic representatives: (1) once in an ancestor of *Cryptosporidium parvum* and (2) a second time in an ancestor of *Plasmodium falciparum*. Therefore, a more detailed analysis of peroxisomal diagnostic markers for biogenesis and metabolism in the Apicomplexa was accomplished in order to examine this observation. Blastp and tblastn searches were conducted in publicly accessible transcriptomic and genomic datasets of other apicomplexan parasites (Table S9/Table S10). In fact, a peroxisome could not be detected by *in silico* analyses in any available apicomplexan parasite belonging to the class of Aconoidasida. Hence, this organelle must have been lost in a common ancestor of this group. Members of the group Conoidasida are more heterogenous: whereas a peroxisome could be detected in coccidian species, the early-branching genus *Cryptosporidium* and the Gregarinasina seem to be peroxisome-less at first sight. However, in the only 20% sequenced genome of *Ascogregarina taiwanensis* (only 6.15 of 30 MB) three crucial peroxins (Pex1, Pex5, Pex7) and two peroxisome-specific PMPs (PMP34 and PMP70) could be identified as well as some metabolic genes (Table S9/Table S10), which indicate that a peroxisome seems to be present in Gregarinasina.

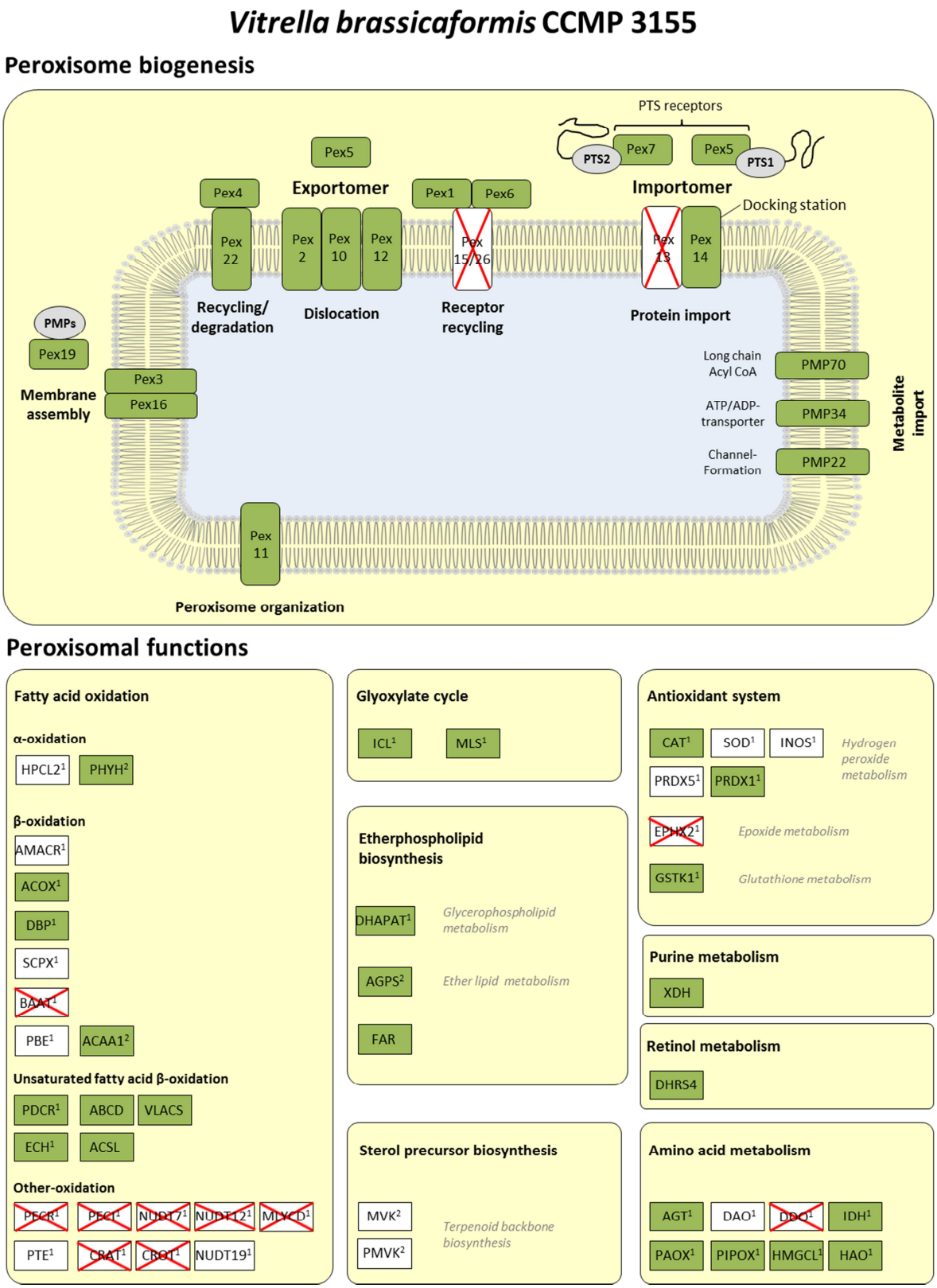


Figure 34: Revised KEGG map for peroxisomal biogenesis and functions. Shown is the presence/absence of peroxins and of metabolic enzymes e.g. for the glyoxylate cycle. The map is exemplary illustrated for the photoautotrophic apicomplexan species *V. brassicaformis*. Presence of proteins is highlighted in green, absence in white. Proteins absent from all alveolate species are crossed in red. Abbreviations: PTS, peroxisomal targeting sequence; Pex, peroxin; PMP, peroxisomal membrane protein.

Furthermore, additional analyses in draft genomes of the dinoflagellates (the suessialean *Symbiodinium minutum*) confirmed the presence of peroxisomes in this alveolate phylum, even in parasitic species (genus *Perkinsus*; Table S11/Table S12). The maintenance of this organelle in the early-branching dinoflagellate *Perkinsus* could be a protection against oxidative stress in this parasite, perhaps predominant in its motile zoospore-stage. In comparison to *P. minimum*, only a few peroxins could be identified in *Symbiodinium*, but perhaps the draft genome of this dinoflagellate species is quite incomplete. Another conspicuous feature is the universal absence of the glyoxylate cycle in apicomplexan parasites (Table S11) that is present in all other peroxisome-harboring alveolates (ciliates, dinoflagellates, chromerids). The fact that the glyoxlyate cycle is even present in the parasitic dinoflagellate genus *Perkinsus* indicates that the loss of this pathway cannot be linked absolutely to parasitism.

4. Discussion

„*Nothing epitomizes the mystery of life more than the spatial organization and dynamics of the cytoplasm*"

(Mitchison 2010)

4.1 Gradual Reduction of Organelle Genomes – Why this Tremendous Gene Loss?

My results revealed that in Apicomplexa and Dinoflagellates (Myzozoa) the genomes of plastids as well as of mitochondria are highly reduced in their genetic content (mitochondria: *Perkinsus*, *Chromera*, *Vitrella*, chapter 3.2). Furthermore, some species seem to have lost the whole organelle genome or even the organelle at all (plastid: *Perkinsus*, *Cryptosporidium*, gregarines, chapter 3.1). This major result of my thesis is comparable to the statement, that in endosymbiotic organelles the driving force seems to be their genome reduction by extensive gene loss (Moran 2002; Diekmann, Pereira-Leal 2013). Genes can get lost if their function is no longer needed. Thus in terms of the endosymbiotic theory a great number of genes, which previously enabled the uptaken bacterium a free-living life style, had disappeared during its transformation into a host-dependent endosymbiont ("*Use It or Lose It*", Moran 2002). On the other hand some already existing nuclear genes can replace organellar genes and *vice versa* in a process designated as "gene substitution". Additionally, genes can be successfully and completely functionally transferred to the nucleus of the host without any counterparts for these genes in the genome of the cell (Martin, Herrmann 1998; Blanchard, Lynch 2000; Timmis et al. 2004). In some cases even dual targeting of proteins to both organelles, mitochondria and chloroplasts, were demonstrated, which hence share a specific function in the most economical way (Peeters, Small 2001).

The reduction of organelle genomes seems to be an ongoing process rather than to be finished. It appears to be an evolutionary advantage to get "centralized" in the cell, like in a good working business, where it is better to have a central authority (nucleus) to manage complex operations with the help of its recruits (compartments). As expressed in Allen et al. (2003): "*One might say that the bipartite or tripartite genetic system of eukaryotic cells is untidy and inefficient.*" In many cases the site of action is sustained by signal sequences for back transfer of proteins into the lumen of the respective organelle. The question arises if gene loss and accordingly transfer is a random or a predestined process. It seems that there

exists a hierarchy of gene loss/transfer: some genes (e.g. of ribosomal origin) are more prone to get lost from the organellar genome than others (e.g. respiratory genes in mitochondria) (Adams, Palmer 2003).

Numerous hypotheses exist of how a gene transfer takes place or is provoked. In the asexual reproducing organelles, one plausible hypothesis is that deleterious mutations can be fixed more frequently (called "Muller's ratchet"), whereupon beneficial mutations can be fixed more rapidly in the nuclear genome. In this case organellar genomes are more prone to genetic drift (change in allele frequency by random variation). Another hypothesis focusses on intra-organellar competition. It seems that "streamlined" organelles have a selective advantage. Furthermore, a possible selective factor for gene transfer from organelles into the nucleus can be the avoidance of free toxic radicals (e.g. reactive oxygen species; ROS) which are produced in the mitochondria and plastids, which enhance the rate of DNA damage in their genome (Blanchard, Lynch 2000; Adams, Palmer 2003). However, peroxisomes are often in spatial vicinity to plastids and mitochondria in order to detoxicate the cell from these free radicals (Lismont et al. 2015; Wanders et al. 2015).

Considering the genomes of the myzozoan organelles, there seems to be indeed an advantage in streamlining the genomic information of the organelle to transfer most of the remaining genes to the nucleus of the protist cell. Transferring this theory in general to single-celled organisms that are constantly exposed to exogenous stress factors to which they must respond quickly, the simplest way would be a system addressing only one compartment, one gene expression machinery and one protein biosynthesis machinery. Thus, this energy-demanding response of the cell would only affect one cellular system.

4.1.1 The Smallest Mitochondrial Genome in *Chromera velia*

The mitochondrial genomes of alveolates seem to experience a strong selective pressure towards the reduction of their genomic content and simultaneously towards the establishment of mitochondrial genes in the nucleus of the host. However, some genes are still encoded in the reduced organellar genomes. This is exemplified by the mitochondrial genomes only harboring three protein-coding genes of the oxidative phosphorylation (OXPHOS; *coxI*, *cox3*, *cob*) in the Myzozoa as well as the discovery of the smallest mitochondrial genome in the chromerid *C. velia* which only harbors the OXPHOS gene *coxI* (Chapter 3.2.1). Why keeping a small part of the organellar genome instead of transferring all into the nucleus of the cell? The CoRR – hypothesis (co-location for redox-regulation) of Allen (1993) assumes that organelles maintain specific genes in order to enable an *in situ* redox regulation of their

expression (Wilson et al. 2012). These genes form the central core components for membrane spanning protein complexes, which are involved in electron transfer and proton translocation. The bioenergetic genes are associated with the environment and are susceptible to abiotic environmental changes. In the case of the respiration processes in mitochondria, oxygen is the major variable input from the environment, in the case of photosynthesis in plastids light and carbon dioxide. If there are changes in the concentration of the respective abiotic factor, the system has to respond to the changes rapidly. The response is realized by direct redox regulation of gene expression and co-location of certain genes and their gene products within the organelle – an inherited trait by their bacterial ancestors (Allen 2003). Additionally, the 'hydrophobicity-' and 'toxicity-' hypotheses are explanations for the retention of certain genes in the mitochondrial genome. In the first case some proteins of the organelle are highly hydrophobic (high mesohydrophobicity level) and cannot be imported easily across membranes, e.g. if the proteins were located in the cytosol, because their genes were transferred to the nucleus, too hydrophobic proteins could not be reimported to their site of action in the lumen of the organelle and are often misrouted to the secretory pathway. Using the example of *coxI* and *cob*, these proteins represent the most hydrophobic proteins in mitochondria (Claros et al. 1995; Heijne 1986). The toxicity-hypothesis implies that the protein products of some genes (*coxI*, *cob*) could be toxic when they are located in the cytosol (Martin, Schnarrenberger 1997; Adams, Palmer 2003).

In the case of the small mitochondrial genome of *C. velia*, the *cob* gene got lost and the *coxI* gene is still encoded on the mitochondrion. So far, the toxicity hypothesis for the omnipresent occurrence of *cob* and *coxI* in mitochondrial genomes could not be experimentally verified. However, the hydrophobicity plots of Claros et al. (1995) clearly show that *cob* and *coxI* are the most hydrophobic mitochondrial genes. Thus, the most plausible explanation for the retention of the *coxI* gene in the mitochondrion of *Chromera* would be the hydrophobicity hypothesis. As the CoRR hypothesis in mitochondria was partially supported in isolated pea leaf mitochondria (Allen et al. 1995), the retention of the *coxI* gene might be also linked to the direct redox regulation of its expression. On the other hand, all other genes of the cell respiration complex were successfully transferred to the nucleus of the cell (Table S7), thus the CoRR hypothesis does not seem to be the raison d'être for the retention of genes in the mitochondrion. Therefore, the mitochondrial genome of *C. velia* represents supposedly the last stage before the complete loss of the mitochondrial genome (Petersen et al. 2014).

The 'Minicircle-Hypothesis' in C. velia

Positive results of inverse PCR (iPCR) experiments with the mitochondria-encoded *coxI* gene (*C. velia*) indicate a circular conformation of the mitochondrial genome in the early-branching apicomplexan species (Chapter 3.2.1).

The combination of the iPCR results with Illumina scaffolds revealed two variants of a continuous circular mitochondrial genome in *C. velia* with a size around only 2 kb, which would hence represent a 'minicircle' (Chapter 3.2.1). Different sequences acquired by this iPCR experiment in addition to the coding region of *coxI*, might indicate the presence of numerous minicircles with different flanking regions around the protein-coding gene as demonstrated in Figure 35.

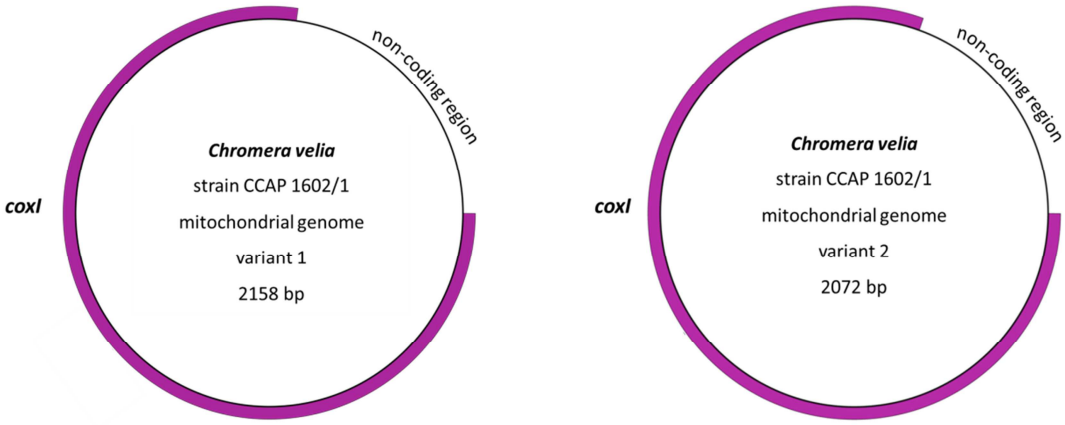


Figure 35: Two variants of the circular mitochondrial genome of *C.velia* inferred by the combination of Illumina nodes and two independent iPCR experiments. The protein-coding region of *coxI* is shown in magenta. The deduced size of this minicircle is 2,158 bp and 2,072 bp, respectively. The visualization of the circular molecule was carried out with OgDraw (Lohse et al. 2013).

Additionally, my conclusions about the mitochondrial genome architecture in the chromerid *C. velia* are contrary to the assumptions of Flegontov et al. (2015). In this publication the authors argue for a linear structure of the *C. velia* mitochondrial genome based on rather fuzzy, not well-resolved Southern blots of for the maintenance of genomic stability topoisomerase-treated genomic DNA with a *coxI* probe (suppl. Fig. S 6 in Flegontov et al. 2015). Furthermore, electron micrographs of mitochondria-enriched DNA by a CsCl-gradient showed numerous linear DNA molecules (Fig. 2 in Flegontov et al. 2015). Accordingly, a hypothetical linear structure of the *C. velia* mitochondrial genome would also be compatible

with the results of my studies. Positive iPCR results might also be obtained if linear molecules produce temporarily circular structures due to fitting ends, e.g. terminal inverted repeats. An alternative explanation is that such fitting ends result in concatenated head-to-tail tandem arrays of linear molecules pretending circularity by positive iPCRs with primer pairs in the coding region of *coxI*. However, I could not find terminal inverted repeats neither within the iPCR results, nor in the Illumina scaffolds. On the contrary, the results of Flegontov et al. (2015) could be also misinterpreted in terms of sheared genomic DNA by harsh treatment. Nevertheless a linear structure of the mitochondrial genome in *C. velia* cannot be excluded with complete certainty and further investigations and experiments are necessary to address this question.

The minicircle hypothesis is everything but digressive. In the euglenozoan genus *Diplonema*, the mitochondrial genomic architecture is very extraordinary (Marande et al. 2005; Marande, Burger 2007). Mitochondrial protein-coding genes are split into four to ten modules per gene which form a unique cassette with flanking regions on a genomic minicircle. This splitting results in more than 100 mitochondrial circular chromosomes classified into class A (6 kb) and class B (7 kb) chromosomes. For example, the *coxI* gene is split up to nine modules, each on a single chromosome, with 7 modules on class A chromosomes and two modules on class B chromosomes. These single modules are joined to a continuous RNA after transcription of the minicircles and the RNAs subsequently undergo trans-splicing and other RNA editing processes (Marande, Burger 2007). This example again demonstrates the plasticity of organelle genomes in protist lineages according to an 'All bets are off' – principle. Plastid minicircles with one to numerous genes per genomic circle are typical for core dinoflagellates (Chapter 1.4.5) and mitochondrial minicircles might reflect an analogous fragmentation of the organelle genome at least in *C. velia*.

4.1.2 A 'Cryptic Plastid' in *Perkinsus olseni*

The record of 'cryptic plastids' is not as frequently as 'cryptic mitochondria'. Plastids in non-photosynthetic lineages are often termed as 'cryptic', because the original function of these organelles, the photosynthesis, got lost (reviewed in Keeling 2010), whereas in 'cryptic mitochondria' such as mitosomes and hydrogenosomes the whole mitochondrial genome is missing (Chapter 1.1.2). Until now the final evidence of the presence of a plastid in the dinoflagellate genus *Perkinsus* sp. is lacking, although it is assumed that this parasite still harbor this organelle (Stelter et al. 2007; Grauvogel et al. 2007; Teles-Grilo et al. 2007; Matsuzaki et al. 2008; Joseph et al. 2010; Fernandez Robledo et al. 2011). In my opinion, this

organelle must be present if there are nucleus-encoded plastid genes containing signal peptides for their relocation to the respective organelle – at least a membrane environed structure where the relict proteins can be transferred back to. Therefore, I could show the presence of the complete plastid-specific MEP pathway for isoprenoid biosynthesis in *Perkinsus olseni* incorporating clearly defined bipartite targeting sequences, which were consistent with previous described dinoflagellate classes of transit peptides (Chapter 3.1.4; Patron et al. 2005). Furthermore, the complete plastid system for Fe-S cluster biogenesis, the Suf system, could be identified likewise containing bipartite signal sequences for a potential transport of these plastid matrix proteins (Table S5).

A DNA-staining approach with the fluorescent stain DAPI in *P. marinus* (Matsuzaki et al. in 2008) was able to visualize the DNA in the nucleus and also in mitochondria, but failed to detect any DNA in potential plastid regions, which were co-localized with the fluorescence-labelled DXR protein of the plastidial MEP pathway. The absence of the DAPI stain could be considered as a negative evidence of plastome absence in a putative plastid of the genus *Perkinsus*. In this study, the genome- and coverage-based analyses in the high quality genomic dataset of *P. olseni* unequivocally show that this dinoflagellate parasite has lost its plastid genome and has transferred all essential plastid genes to the nucleus (Chapter 3.1.4, Table S5). Unfortunately, the complementary ultrastructural evidence of the plastid is still missing. However, we are in the golden age of sequencing techniques, which allows us to analyze whole genomes and to establish high amounts of sequence information. With all these data the genetic secrets of organisms get transparent and allow us to formulate substantial hypotheses as well as to draw meaningful conclusions. Michael L. Ginger already realized in 2006 that '[…] *genome sequencing provides a powerful tool for ascribing function to organelles that are otherwise refractory to experimental study.*'

Plastid Loss in Alveolate Species

One can assume that a plastid without a genome is the pre-stage of plastid loss. Accordingly, I compared the plastid functions of *Perkinsus* with closely related species that actually lost their plastid organelle. For example the apicomplexan genus *Cryptosporidium* apparently lacks a plastid and contains a mitochondrion-related organelle, the mitosome (Riordan et al. 1999; Abrahamsen et al. 2004; Henriquez et al. 2005). A phylogenomic study proposed that *Cryptosporidium* evolved from a plastid-containing lineage and that it has lost its apicoplast during evolution, but this prediction was not confirmed until now (Huang et al. 2004).

For a long time, the exact phylogenetic position of *Cryptosporidium* as an early-branching apicomplexan genus together with the other deep-branching apicomplexan lineages, chromerids and gregarines, was not clearly resolved. Because of the limited availability of whole genome sequences of these lineages, multiple apicomplexan phylogenies only based on single genes or morphology data. In the following section I outline phylogenies representing the host cell (nuclear markers) in order to address the question of the basal topology in Apicomplexa. In 2004, a phylogenetic study based on two single gene phylogenies of a nuclear heat shock protein (hsp90) and the actin gene incorporated *C. parvum* and the gregarine species *Monocystis agilis*. This study resulted in two different positions of these early-branching apicomplexan species as either a monophyletic sister group or as serial branching sisters to the other apicomplexan species (Leander, Keeling 2004). Both, *Cryptosporidium* and gregarines, are assumed to lack a plastid (Chapter 1.4.5). Both phylogenies show the aplastidial apicomplexan lineages as a sister to the apicoplast-bearing lineages.

The discovery of the two photoautotrophic algae ("chromerids"), *C. velia* and *V. brassicaformis*, which are close relatives to the parasitic non-photosynthetic Apicomplexa made the discussion about alveolate plastid evolution more complex (Moore et al. 2008; Obornik et al. 2012). Different analyses indicated a sister relationship of these two algal species to the apicoplast-bearing species by diverse topologies (monophyletic vs. serial branching; Janouskovec et al. 2010; Obornik et al. 2012).

The first phylogenetic study integrating species of all three early-branching apicomplexan lineages (chromerids, *Cryptosporidium*, gregarines) is based on 73 ribosomal proteins and more than 15,000 amino acid positions (Bachvaroff et al. 2014). This ribosomal tree shows *C. velia* as the deepest branch of the Apicomplexa, followed by the gregarine species *Gregarina niphandrodes*. The monophyletic group of the genus *Cryptosporidium* is the sister group to all apicoplast-bearing parasites in this study.

However, almost contemporaneously to our phylogenomic studies, a Czech working group (Janouskovec et al. 2015) published a species tree including a broad alveolate taxon sampling also including the non-photosynthetic colpodellids. In this tree, the chromerids *C. velia* and *V. brassicaformis* branch among the non-photosynthetic colpodellids forming the sister group 'chrompodellids' to all other apicomplexan lineages. The aplastidial genus *Cryptosporidium* is the sister group to the apicoplast-bearing parasites. The study of Janouskovec et al. (2015) and also the study of Bachvaroff et al. (2014) corresponds widely to our phylogenomic tree (chapter 3.1.1, Fig. 4). In this tree the monophyletic chromerids form the earliest branching

apicomplexan lineage followed by *Cryptosporidium* spp. as a sister to the plastid-bearing parasites. Additionally, we introduced the gregarines to this tree in exchange to the colpodellids, because they were initially irrelevant for the issue of a plastid loss in *Crypotsporidium* and gregarines. In our phylogenomic analysis the gregarines form the sister to the genus *Cryptoporidium.* With a bootstrap value of 76% the position of *Cryptosporidium* spp. is not well resolved and a monophyletic grouping of both aplastidial lineages is still possible.

I summarized the existing phylogenomic information in two distinct scenarios of how plastids could be recruited among apicomplexan lineages considering our species tree (Fig. 4) as the reference. The most parsimonious assumption would be that the plastid was recruited in a common ancestor of all extent apicomplexan lineages, including plastid loss in aplastdial lineages (Figure 36, scenario A). The branching order of *Cryptosporidium* and gregarines is not resolved by previous phylogenetic analyses (Fig. 4), thus the amount of plastid losses is one point of interest in this scenario (Figure 36, A (1)-(2)). Plastid losses can be opposed to multiple plastid gains (Figure 36, scenario B). The scenario B hypothesizes that the common ancestor of all extent species did not harbor a plastid organelle. Consequentially, the plastid-bearing lineages must have acquired their plastids secondarily and independently: in a common ancestor of chromerids as well as in a common ancestor of all plastidial apicomplexan parasites. Again, both possible topologies of *Cryptosporidium* and gregarines are shown (Figure 36, B (1)-(2)).

The current PhD thesis confirmed the hypothesis of a secondary plastid loss in *Cryptosporidium* spp. and gregarines with the phylogenomic analysis of the host cell in comparison to single gene phylogenies of nuclear-encoded plastid markers. Plastome-encoded markers were discarded due to their fast evolving behavior in *Chromera* and *Vitrella*. The analyses first confirmed the basal positioning of the photosynthetic chromerids as the earliest branching apicomplexan lineage (Figure 36, scenario A) and plastid markers moreover documented the monophyly of Apicomplexa. Based on this result, the loss of the apicoplast must have occurred at least once. In our phylogenomic study Scenario A (1) with two independent plastid losses (Figure 36) is favoured due to the serial branching order of gregarines and Cryptosporidium to the apicoplast-bearing parasites (Fig. 4). Thus, this comparative study of host cell and plastid-specific markers represents the first clear evidence of plastid loss in aplastidic apicomplexan lineages.

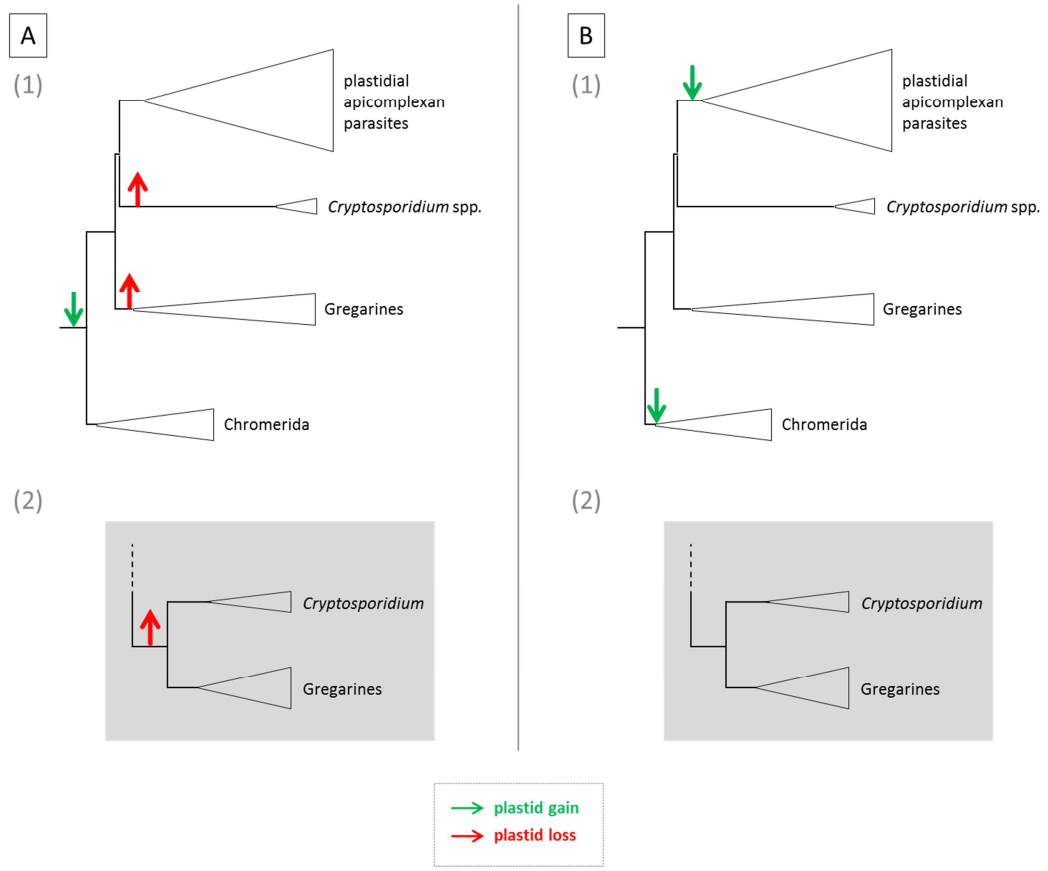


Figure 36: Two different scenarios of how plastids were acquired among apicomplexan lineages. Plastid gain is marked by a green arrow, plastid loss by a red arrow. A) Most parsimonious scenario of a common plastid gain in all apicomplexan lineages implying plastid loss in aplastidial lineages with either (1) two losses in separate ancestors of *Cryptosporidium* spp. and gregarines or (2) one loss in a common ancestor. B) Two independent plastid gains in plastid-bearing lineages: in a common ancestor of chromerids as well as in a common ancestor of the plastidial apicomplexan parasites. Scenario B implies that *Cryptosporidium* spp. and gregarines never harbored a plastid. Two possible tree positions for these two aplastidial lineages are shown in (1) and (2) corresponding to Scenario A.

In contrast to *Perkinsus*, *Cryptosporidium* seems to have overcome the plastid dependency by maintaining the cytosolic pathway for fatty acid synthesis (FAS I) and redirecting host isoprene precursors and tetrapyrrole heme for its own benefit (Zhu et al. 2000; Xu et al. 2004; Bessoff et al. 2013; Gornik et al. 2015). Additionally, the sole function of the mitosome seems to be the synthesis of Fe-S clusters by the Isc system that can probably replace the Suf pathway of a plastid in combination with the still present cytosolic iron-sulfur assembly system in this genus (LaGier et al. 2003; Tsaousis et al. 2014). Free-living cell stages (oocysts) of this parasite challenge the organism in absorbing crucial metabolites from its hosts such as the vital isoprenoid precursors, thus *Cryptosporidium* must have developed a technique to survive 'metabolic starving conditions'.

In a recent study of another marine parasite of crustaceans, the dinoflagellate *Hematodinium* sp., it was clearly demonstrated that the organism is viable without a plastid by the retention of ancestral anabolic pathways, enzyme relocation from the organelle to the cytosol and

finally by scavenging of metabolites from its host. This process was considered as an 'endosymbiosis undone' because *Hematodinium* sp. managed to stepwise eliminate its plastid and to live independently of its former endosymbiont (Gornik et al. 2015). This dinoflagellate seems to be an example of the ability to adapt perfectly to its parasitic life style in order to reduce completely its own energy demanding metabolism. Although the complete life cycle of *Hematodinium* sp. is still not entirely resolved, in addition to the internal parasitic cells in crustacean hemolymphs, free living dinospores as a waterborne transmission agent were detected (Frischer et al. 2006).

An explanation, of how free living stages of both parasites could survive without crucial metabolites, is the fast exploration of suitable alternate hosts by these parasites. Such a possibility was already described for *Cryptosporidium parvum*, whose oocysts could infect oysters in estuarine waters, which represent putative mechanical vectors for their actual hosts, mammals. It could be shown that the uptaken oocysts are still infectious for mice after retaining them in oyster tissues for one week (Fayer et al. 1998).

However, it is known that the life cycle of *Perkinsus* sp. comprises also a free living motile cell stage in the marine water and a parasitic cell stage in the internal milieu of the crustacean (Fernandez Robledo et al. 2011). *Perkinsus* seem to have adapted perfectly to its two cell stages maintaining the most important plastidial pathways (MEP and SUF; Yeh, DeRisi 2011; Gisselberg et al. 2013) as well as complete functional mitochondria (Chapter 3.1.4; Chapter 3.2.2). The key to success is here the streamlining of their organelles by the transfer of their genetic content into the nucleus and by maintaining the compartmentalization of some metabolic reactions in the cell instead of abandon them and reconstruct an 'open system' in the cytosol. In contrast to 'endosymbiosis undone' I would classify this process as 'endosymbiosis accomplished'.

The 'Plastid Without Plastome'-Hypothesis

In the first instance the hypothesis of a plastid organelle without a genome sounds adventurous, but it does not seem to be an individual case in *Perkinsus* spp.. The same was already described in the obligate parasitic plant *Rafflesia lagascae* using a similar approach like in this study. The genome of the parasite was sequenced with ~ 350 x depth coverage, but there were no detectable essential genes for a plastid genome in the presence of a complete reconstructed mitochondrial genome (Molina et al. 2014). The authors of this study stated that this putative plastome loss is possibly linked with the very ancient origin of parasitism in this family of plant parasites (Xi et al. 2013). Their 'plastid without a plastome'-hypothesis is

supported by ultrastructural transmission electron micrographs depicting plastid-like compartments with homogenous stroma (Molina et al. 2014).

Another example for a possible plastid organelle without a genome is the green algal non-photosynthetic genus *Polytomella*. Again, this approach uses the high throughput sequencing technique in order to demonstrate the absence of plastid-derived sequencing reads (Smith, Lee 2014). Additionally, discarding genes for photosynthesis, energy production and the expression apparatus, they compared the plastomes of the closely related *Chlamydomonas reinhardtii* and the likewise non-photosynthetic green alga *Helicosporidium* with their genomic sequence data of *Polytomella*. All remaining hypothetically essential plastid-encoded genes could be either shown to be nuclear encoded, functionally modified or they got lost.

Cases of "cryptic plastids" comparable to *Perkinsus* are the heterotrophic dinoflagellates *Oxyrrhis marina* and *Crypthecodinium cohnii*. In both early-diverging species, the presence of the plastid organelle is disputed, because neither a plastid genome nor a plastid structure is detectable. Anyway, plastid-derived genes encoded on the nucleus were demonstrated in EST datasets in both species (Sanchez Puerta et al. 2007; Slamovits, Keeling 2008). These identified plastid-derived genes indicate that these early-branching heterotrophic dinoflagellates evolved from a photosynthetic ancestor and also maintained this organelle. The idea emerges, that in these close relatives of *Perkinsus*, a hypothetical plastid was also reduced to a 'reaction chamber' for essential metabolic functions. Anyway, *Oxyrrhis* and *Crypthecodinium* are marine 'hunters' feeding on numerous protists (Hansen et al. 1996; Ucko et al. 1997), whereupon they could possibly steel genetic material from the plastids of their prey without harboring these organelles by themselves and perhaps modify these proteins in their function.

Each example discussed here advocates the possibility of the 'plastid without plastome' hypothesis and thus justifying the term 'Perkinsuplast' for such plastids in *Perkinsus* species (Chapter 3.1.4; Grauvogel et al. 2007).

4.2 Reassessment of the 'Peroxisome Hypothesis' in Alveolates

4.2.1 Presence of the 'Neglected Organelle' in Myzozoan Key Species

In alveolates the term 'peroxisome' was initially introduced in the ciliate *T. pyriformis* (Duve, Baudhuin 1966), whereas in the remaining two alveolate phyla summarized to the group of 'Myzozoa' (Cavalier-Smith, Chao 2004) the absence of the peroxisome organelle was assumed (Schlüter et al. 2006; Gabaldon 2010). The parasitic group of Apicomplexa was even

declared as "[…] *the first group of organisms devoid of peroxisomes, in the presence of mitochondria.*" (Schlüter et al. 2006).

By means of an exhaustive search in whole genomic and transcriptomic datasets of myzozoan key organisms using peroxisomal markers inferred from the KEGG Pathway Map (Kanehisa, Goto 2000), I was able to detect the presence of crucial peroxisomal proteins in apicomplexan and dinoflagellate species (Chapter 3.3.1, Table S9 – Table S12). In the newly established transcriptomes of both free-living photoautotrophic chromerids (*C. velia*, *V. brassicaformis*), which are the closest relatives of the apicomplexan parasites, a large number of peroxisomal markers was discovered with the highest amount of detected peroxisomal markers in *V. brassicaformis* (Figure 34), which will be therefore well suited as a model organism for further peroxisomal functional studies in Alveolata. Surprisingly, a fully functional peroxisome organelle seems to be universally present in the parasitic apicomplexan class Conoidasida (*Toxoplasma*, *Hammondia*, *Neospora*, *Sarcocystis*, *Cyclospora*; Table S9 – Table S12). The detection of peroxisomes in this apicomplexan class disproves previous assumptions about the general absence of this organelle in *Toxoplasma* (Ding et al. 2000; Gabaldon 2010), although in a recent publication this question was addressed again (Gabaldon et al. 2016).

In dinoflagellate datasets of four considered species representing photoautotrophic as well as parasitic members (*Prorocentrum*, *Symbiodinium*, *Perkinsus* spp.; Table S11 – Table S12), the presence of the peroxisome organelle could be demonstrated. However, the genome sequencing of the large dinoflagellate genomes is in general challenging demonstrated by the partial draft genome of *S. minutum*, in which so far only three peroxins (Pex1, Pex5, Pex12) could be detected (Table S11). Because of its 'incomplete' genomic dataset, it is possible that more peroxins are actually present in *S. minutum.* In contrast a broad spectrum of metabolic markers could be identified in this partial genome, additionally indicating the presence of the peroxisome (Table S12). The function of a peroxisome in parasitic representatives of alveolates (Conoidasida, *Perkinsus*) could be the protection against oxidative stress either during the penetration of the parasite into its host (Conoidasida, *Perkinsus*) or during free-living motile stages (*Perkinsus*) when the organism is prone to abiotic factors (Bosch et al. 2015).

This study provides the first detailed *in silico* evidence for a peroxisome organelle in myzozoan species. The observation that peroxisomes are present in all three alveolate phyla indicates that this organelle was already present in a common ancestor of alveolates. The common peroxisomal ancestry is supported by phylogenetic analyses of three markers in this

study (Figure 30, Figure S23). Due to the current investigations, ten peroxins serve as diagnostic markers to unequivocally detect the presence of a peroxisome in alveolates based on available genome sequences (Pex1, Pex2, Pex5, Pex6, Pex7, Pex11, Pex12, Pex16, PMP34, PMP70; Table S9 – Table S12).

4.2.2 Independent Secondary Losses of Peroxisomes in Parasitic Apicomplexa

According to the detection of a peroxisome organelle in the parasitic group of Conoidasida, I could otherwise demonstrate the absence of all but two peroxisomal markers in two parasitic lineages of the Apicomplexa (Aconoidasida, *Cryptosporidium*; chapter 3.3.1, Table S9 – Table S10). Two peroxins (Pex4, Pex22; Table S9) were indeed detected, but these genes could represent paralogs or modified genes, which have adopted other functions within the cell. The absence of all peroxisomal diagnostic markers within the Aconoidasida (*Plasmodium*, *Babesia*, *Theileria*) and the early-branching apicomplexan genus *Cryptosporidium* indicates that the organelle was lost twice in a common ancestor of each lineage given the common peroxisomal ancestry in all alveolates (Chapter 3.3.1). These parasites could have abandoned this system perhaps due to a high degree of specific adaptation to their respective hosts.

4.3 Genetic Mosaicism in Alveolate Genomes

My phylogenetic analyses and blast search results showed that horizontal gene transfer occur more often than previously expected. Horizontal gene transfer could be classified into genes with a xenologous origin and genes, which were acquired by the endosymbiont (endosymbiotic gene transfer, EGT). In general, if some phylogeny patterns occur regularly the hypothesis arises that this observation is not a random phenomenon and might reflect a common evolutionary history. Numerous genes are of cyanobacterial origin in complex algae (e.g. TIC20, SUFB; Chapter 3.1.2/3.1.5) indicating that the plastid organelles were once established by the uptake of a cyanobacterium. ‘Green algal’ and ‘red algal’ (DXS, CDC48; Chapter 3.1.2) signals in genomes of complex algae indicate that a number of organelle genes was once transferred from the compartment to the nucleus of the host cell thus reinforcing the bond between endosymbiont and host.

In individual cases xenologous sequences of bacterial origin were established in the nuclear genome of the host without an endosymbiotic background. For instance in the DXS phylogeny of this study, it appears that the plastid gene was once replaced by an α-proteobacterial homolog before the secondary endosymbiosis took place (Figure S5).

Additionally, the *hdr* gene of the plastidial MEP pathway was once replaced by chlamydial bacteria exclusively in apicomplexan species (Petersen et al. 2014; Figure 14; Figure S6).

4.3.1 'Chlamydial Genes' in Alveolates

Frequently, a phylogenetic connection between complex alga of interest and the bacterial order chlamydiales is found. In algal phylogenies it is noticeable that some plastidial genes are of chlamydial origin, e.g. observed for the nuclear-encoded MEP genes MCT and CMK, as well as shown for HDS and HDR in this study (chapters 3.1.2). Because of the frequency of this phylogenetic connection, an ancient HGT of chlamydiales to the ancestor of photosynthetic eukaryotes was already discussed (Huang, Gogarten 2007; Becker et al. 2008). In 2007, Huang and Gogarten inferred that this high number of transferred genes is rather the product of an ancient chlamydial endosymbiosis with the ancestral primary photosynthetic eukaryote than a random shift of genes by HGT. Furthermore, they argue that this ancestral endosymbiotic relationship facilitated the establishment and rise of the primary plastids, which is presented by "ménage à trois" in the publication of Facchinelli et al. (2013): a tripartite symbiosis consisting of the cyanobiont, the chlamydial bacterium and the host. The chlamydiales are obligate intracellular bacteria found in a number of organisms, excluding photosynthetic members. Thus, I agree with Becker et al. (2008) that perhaps a close relationship between chlamydial bacteria and an ancestor of photosynthetic algae existed and that accordingly ancient horizontal gene transfers occurred, because of the abundance and persistence of these bacteria in early eukaryotes.

The phylogenetic analysis of the *hdr* gene shows that all apicomplexan sequences including the *C. velia* sequence have a chlamydial origin, whereas the sequences of all other complex algae are of cyanobacterial origin (Fig. 5 in Petersen et el. 2014). The phenomenon of a chlamydial HGT only in apicomplexan species (Figure 14/Figure S6 in this study) could have two possible explanations: (1) this HGT represents a very ancient event given that the ancestor of all algal/parasitic lineages once incorporated two *hdr* copies, one from the cyanobiont and one from a chlamydial bacterium. The ancestor of all apicomplexan species possibly lost the cyanobiont copy and only retained the chlamydial copy, whereas the other algal lineages retained only the cyanobiont copy; (2) this HGT represents a more recent event that a common ancestor of all apicomplexan lineages once acquired this gene by chlamydial bacteria and consequently lost its cyanobiont copy. However, this example clearly demonstrates that the chlamydial *hdr* gene was once acquired by a plastidial ancestor of modern Apicomplexa via HGT. The grouping of all apicomplexan parasites with the

photoautotrophic alga Chromera proves the common origin of their plastids by higher order endosymbiosis.

4.3.2 A 'Green Footprint' in *Chromera velia*

C. velia represents a complex alga of red algal origin (Janouskovec et al. 2010). Nevertheless, several authentic 'green' genes could be identified in this alga (Woehle et al. 2011; Burki et al. 2012). Accordingly, in this thesis a close relationship between the SufB protein of *C. velia* and that of the green lineage was described, especially the surprisingly high number of shared intron positions between SufB of *C. velia* and the green alga *C. reinhardtii*. Comparing this 'green footprint' with instances of cryptic origins in genes of other protist groups should shed light on this mystery.

In 2006, Petersen et al. already showed that the plastid-located phosphoribulokinase (PRK) of the Calvin cycle has a green affinity in members of the CASH group (complex red lineage) and they deduced from this result that the *prk* gene must have been recruited by horizontal gene transfer with a green alga (Petersen et al. 2006). Moustafa et al. described in 2009 green algal footprints in the nuclear genome of the diatoms *T. pseudonana* and *P. tricornutum* with numerous genes in common with haptophytes indicating a cryptic secondary endosymbiosis with a green alga in a common ancestor of CASH lineages. Additionally, a recent study of membrane transporters in these diatoms confirmed the green ancestry in the majority of the examined genes (Chan et al. 2012). Both publications are in agreement with the hypothesis that in a common ancestor of CASH lineages a cryptic green algal endosymbiont was once present until this endosymbiont was replaced by a red algal antagonist that represents the contemporary plastid (Dorrell, Smith 2011). The other way round, in the green lineage Chlorarachniophyta multiple genes of red algal origin were identified in nuclear genomes of the two species *Amorphochlora amoeboformis* and *Bigelowiella natans* (Yang et al. 2014). Because these red genes were plastid-related, it was speculated that they originated from photosynthetic ancestors. These findings resulted in the hypothesis that there must have been either a cryptic red algal endosymbiont before the second endosymbiosis with a green alga or a scavenging of red algal genes by long-term feeding with red algal species after the green algal plastid was established. Both cases in diatoms and Chlorarachniophyta could reflect that the host in secondary/tertiary endosymbiotic scenarios was possibly not fastidious in the assimilation of prey organisms and that the heterotrophic situation was very dynamic also in the exchange of genetic material.

In contrast to a complex scenario that bases on a cryptic endosymbiosis, the evolutionary history of the SufB protein in *C. velia* can be explained comparable to the 'green' PRK in members of the CASH group (Petersen et al. 2006). The Suf system for iron-sulfur cluster assembly represents one of the most ancient biogenesis pathways (Chapter 1.5.2). The fact that the *sufB* gene is nuclear-encoded in *C. velia* in contrast to all other plastid-encoded *sufB* sequences in alveolates, indicates that *C. velia* once acquired its *sufB* sequence from a green algal ancestor by horizontal gene transfer to its nucleus. Consequently, the possibly former plastidial copy was lost during evolution because of the abundance of *sufB* copies in the cell. The exceptional number of shared intron positions between *C. velia* and *C. reinhardtii* (Figure S12) suggests that the *sufB* sequence came from an ancestor of *Chlamydomonas*.

4.3.3 Gain and Loss of Metabolic Genes in Peroxisomes

Phylogenetic analyses of peroxisomal metabolic proteins indicate that the peroxisomal metabolome demonstrates "*a large degree of mosaicism*" provoked by extensive horizontal gene transfer which results in "*mosaic pathways consisting of genes acquired from many different sources.*" (Schnarrenberger, Martin 2002). The third considered organelle of this study seems to be the most genetically dynamic compartment responding to the special needs of the respective organism and even of different cell stages or tissue types (Chapter 1.3). Such a 'mosaic nature' (Gray 2015) often described for mitochondrial genes, is clearly demonstrated by the three subanalyses of the glyoxylate protein malate synthase (MLS, chapter 3.3.2, Figure 32), which suggest different independent origins for this gene in the three alveolate phyla (MLS-1: unresolved bacterial origin; MLS-2: γ – proteobacterial origin; MLS-3: eukaryotic origin). This pattern of xenologous gene replacement could be explained by iterative loss and regain of the glyoxylate cycle in alveolates. This high degree of HGT is not a complete novelty: it was often found that a significant portion (13 – 18%) of the peroxisomal proteome is of α-proteobacterial origin indicating the recruitment of mitochondrial counterparts such as enzymes of the fatty acid metabolism (Gabaldon et al. 2006; Tabak et al. 2006). Additionally to the α-proteobacterial origin, Gabaldon et al. (2006) found frequently the association of peroxisomal metabolic proteins with diverse other bacterial groups demonstrating that "*[…] such diverse origins underscores the ease at which the peroxisomal proteome can recruit new proteins*". Thus these patterns highlight the plasticity of the peroxisomal proteome. It is likely that during evolution some lineages lost metabolic functions because they have either been replaced by functions of other organelles such as mitochondria, by metabolites scavenged from a host or because they do not need this

function any more due to a changing life style. Later on, some lineages could have regained such metabolic functions by recruiting similar nuclear-encoded proteins, which are double targeted to mitochondria and peroxisomes or by recruiting bacterial homologs via horizontal gene transfer.

4.4 Summary and Outlook

This work highlighted the highest degree of genomic reduction in alveolate organelles: (1) the smallest mitochondrial genome encoding only one protein-coding *coxI* gene in *C. velia*, (2) the loss of the plastome with simultaneously retaining a putative plastid structure in *P. olseni*, (3) culminating in the secondary loss of the entire plastid organelle in the genus *Cryptosporidium* and in gregarines. The reason of this immense 'reductive evolution' is still not elucidated, but the process itself was declared as "*[…] the dominant mode of evolution*" (Wolf, Koonin 2013).

The presence of the smallest mitochondrial genome ever found in the photoautotrophic alga *Chromera* demonstrates that genome reductions are not only reserved for parasitic organisms. Its closest relative, the photoautotrophic alga *V. brassicaformis*, also harbors a strong reduced mitochondrial genome encoding for the two proteins COXI and COB, whose genetic code is fused to one combined open reading frame. However, the structure of both mitochondrial genomes seems to be different. We could show via the combination of iPCR and Illumina data that there are indications of mitochondrial genomic minicircles in *C. velia*, whereas in *V. brassicaformis* a linear genomic conformation is assumed due to failed iPCR experiments. The final proof for the genomic structure of chromerid mitochondria is still missing, thus Southern Blot analyses would be the following method of choice in order to investigate this unanswered question. Previous Southern Blots were accomplished for *C. velia* via restriction digestion of its genomic DNA and a digoxigenin-labeled *coxI* probe. For instance, the EcoRI restriction enzyme cuts three times within the *coxI* gene. If the mitochondrial genome of *C. velia* is circular, accordingly three fragments should appear on the Blot by hybridization with the *coxI*-probe. In the case of a possible linearity of the mitochondrial genomic molecule, four *coxI* fragments should be identified. Unfortunately, the first Southern Blots provided only conflicting results and weak signals for fragments. It is possible that the use of the whole extracted genomic DNA of *C. velia* is to blame for the malfunctioning experiments. Thus, the extraction of only mitochondrial genomic DNA by cell fractionation on a cesium chloride gradient or by a commercial Kit could lead to results with higher resolution in experiments,

possibly also in tailing PCR experiments, which were proven to be challenging in first tests with *V. brassicaformis* and *P. olseni* genomic DNA. Southern Blots with extracted mitochondrial DNA should also be applied to the parasitic dinoflagellate *P. olseni* in order to investigate its mitochondrial genomic structure. In this organism the two mitochondrial protein-coding genes *coxI* and *cob* could be identified, but could not be shown to be located on one single genomic molecule in this study. Southern Blot experiments could demonstrate if these genes are encoded on two different mitochondrial molecules or actually on one single molecule. Furthermore, the Southern Blot would simultaneously give information about a linear or circular conformation of the mitochondrial genome/s.

The discovery of nuclear-encoded plastid genes with well characterized bipartite signal sequences, as well as the absence of plastid-encoded genes (e.g. 16S rRNA) in *P. olseni* results in our hypothesis of a plastid without a plastome, namely the 'Perkinsuplast' (Grauvogel et al. 2007). This hypothesis has to be proven by further ultrastructural experiments in order to identify a plastid structure to which the nuclear-encoded plastid-targeted proteins can be transported. As our immunogold experiments were unsuccessful, this method has to be improved. In particular, the handling of the zoopores in the preparation step has to be corrected in order to visualize the accurate cell stage. In parallel it would be of interest if other supposed 'Perkinsuplasts' exist, for instance in the two heterotrophic dinoflagellates *O. marina* and *C. cohnii*, which also encode for plastid-derived genes in their genomes. Therefore, the development of immunogold experiments for these two heterotrophic non-photosynthetic dinoflagellates and perhaps further non-photosynthetic relatives would shed light on the distribution and frequency of such 'Perkinsuplasts' in combination with the *in silico* detection of nuclear-encoded plastid markers.

This study excluded the group of colpodellids comprising free-living non-photosynthetic predators, but which were shown to form a group with the photoautotrophic chromerids on a phylogenomic level called the 'Chrompodellids' (Janouskovec et al. 2015) as a sister to all other apicomplexan lineages. A transcriptomic analysis of the colpodellid *Voromonas pontica* identified markers for plastid-associated pathways, which partly carried plastidial signal and transit-like signal sequences (Gile, Slamovits 2014). Thus, it would be of interest if these free-living relatives of apicomplexans also harbor a 'cryptic plastid' like shown for *P. olseni* or if they rather recruited such plastid-derived genes by their prey. Therefore, analyses of whole genomic datasets of colpodellids as well as ultrastructural approaches should be targeted.

Wolf and Koonin (2013) explained that there "*[…] is not a one-way path of reduction*" and that "*[…] the reduction ratchet is constrained by the advantages of retaining certain*

metabolic pathways that complement the host metabolism". Thus, the composition of organellar genomes and their degree of reduction or rather the transfer rate of their genetic content to the nucleus is probably dependent on a complex combination of abiotic and biotic factors. This plasticity of organelles is clearly demonstrated by the peroxisomes which can obviously lose and regain metabolic functions with regard to the occupied niche of the respective organism. In order to complete the inventory of peroxisomes in available whole genomic datasets of alveolates, the list should be upgraded continuously for peroxisomal diagnostic markers in those species.

5. Supplemental Material

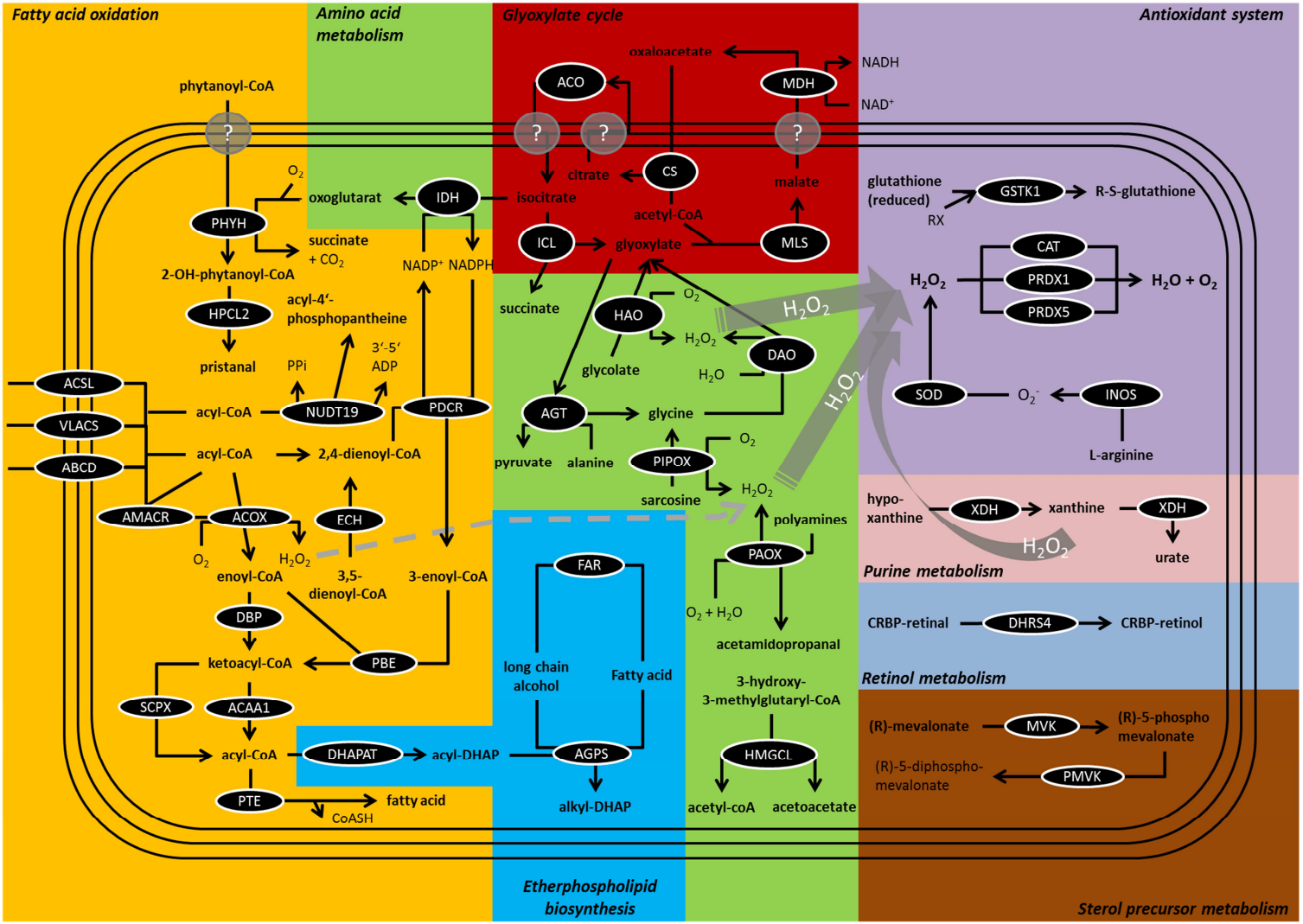


Figure S1: Schematic overview of metabolic pathways in the peroxisome organelle. Each pathway is assigned to a distinct color. The most important involved proteins are encased in black. The directions of reactions are characterized by black arrows. Grey arrows symbolize the release of toxic hydrogen peroxide by some peroxisomal enzymes. Grey circles with question marks represent transporters which are until now unknown. Pathways and metabolites inferred by the Peroxisome Map of KEGG (Kanehisa, Goto 2000) and the Metabolome Map of the PeroxisomeDB 2.0 (Schlüter et al. 2010, 2010).

```
Venn.dat <- read.csv("File location/File_name.csv",
header=TRUE, sep=";", fill=TRUE, quote="\"")
names(Venn.dat) <- c("Gene","Cyanophora paradoxa", "Chondrus
crispus", "Arabidopsis thaliana", "Guillardia theta",
"Plasmodium falciparum", "Dinophyta minicircle",
"Thalassiosira pseudonana", "Emiliania huxleyi")
summary(Venn.dat)
m <- data.matrix(Venn.dat[ ,c(2,3,4,5,6,7,8,9)])
require(venneuler)
y <- venneuler(m)
plot(y, col=colID)
summary(y)
colID <- c("turquoise", "red", "green", "pink", "orange",
"yellow", "brown", "violet", alpha=0.3)
legend(0.05, 0.09, legend =
colnames(m),fill=makeTransparent(colID,alpha=0.3),
x="bottomright")


makeTransparent = function(..., alpha=0.5) {

  if(alpha<0 | alpha>1) stop("alpha must be between 0 and 1")

  alpha = floor(255*alpha)
  newColor = col2rgb(col=unlist(list(...)), alpha=FALSE)

  .makeTransparent = function(col, alpha) {
    rgb(red=col[1], green=col[2], blue=col[3], alpha=alpha,
maxColorValue=255)
  }

  newColor = apply(newColor, 2, .makeTransparent, alpha=alpha)

  return(newColor)

}
```

Figure S2: R-script for the reconstruction of e.g. a Venn diagram incorporating plastome data of eight different eukaryotic organisms.

Table S1: Assembly of established primer pairs utilized in conventional PCR experiments of this study. These experiments concentrated on the completion of genes of the plastidial MEP pathway for isoprenoid biosynthesis in the photosynthetic apicomplexan species *C. velia*. Furthermore, cDNA sequences of two MEP proteins were established for the crustacean parasite *P. marinus* for antibody production against these sequences. Additionally, the presence of *Mycoplasma* bacteria often contaminating cell cultures as well as the presence of bacterial/plastidial 16S rDNA was checked with universal PCR primers.

Conventional PCR experiments: Established primer pairs for genes of interest						
#	**Name**	**Sequence**	**Length**	**Gene**	**Organism**	**Tm [°C]**
P164	PmDXR-for	ATATGGATCCGTTAAGCGAGTAGCAGTC	28mer	*dxr*	*P. marinus*	61
P165	PmDXR-rev	ATAAAAGCTTACCAACTCATCAGCAGGA	28mer	*dxr*	*P. marinus*	59
P168	PmCMK-for	ATATGGATCCGCTGTGGTTAGGAGGGCAGGG	31mer	*cmk*	*P. marinus*	70
P169	PmCMK-rev	TGCTAAGCTTCAAACGGAGTCATACAGA	28mer	*cmk*	*P. marinus*	61
P238	16S-9b_for	GRGTTTGATCCTGGCTCAG	19mer	*16S*	universal	56
P239	16S-1406_rev	ACGGGCGGTGTGTRCAA	17mer	*16S*	universal	60
P686	DXR_Cv_for	ACCATCACCTTGCCAGAGAC	20mer	*dxr*	*C. velia*	60
P687	DXR_Cv_2_for	TTCTTTCGGAACCGGAGAC	19mer	*dxr*	*C. velia*	60
P688	DXR_Cv_rev	GAAGTTGATCGTGGTATCGGTAA	23mer	*dxr*	*C. velia*	60
P691	CMK_Cv_for	GAACCCACCCACCTTGTTC	19mer	*cmk*	*C. velia*	60
P692	CMK_Cv_2_for	ACCCACCTTGTTCCACACTT	20mer	*cmk*	*C. velia*	59
P693	CMK_Cv_rev	GCGAAAGGCTTCATTAATTCAC	22mer	*cmk*	*C. velia*	60
P730	Myco 3’	GCGGTGTGTACAARMCCCGA	20mer	*16S*	*Mycoplasma*	62
P731	Myco 5’	YGCCTGRGTAGTAYRYWCGC	20mer	*16S*	*Mycoplasma*	59

Table S2: Assembly of established primers utilized in inverse PCR experiments for the review of a possible circularity of mitochondrial genomes in three different protist species: *C. velia*, *V. brassicaformis* and *P. olseni*.

Inverse PCR experiments: Established primer pairs for mitochondrial genes						
#	**Name**	**Sequence**	**Length**	**Gene**	**Organism**	**Tm [°C]**
P668	Cv_Mito_rev	CCTAAACCAACACCAAAATCT	21mer	*coxI*	*C. velia*	51
P670	Cv_Mito_rev	GATAAAAATAATAAATGAAAA	21mer	*coxI*	*C. velia*	36
P671	Cv_Mito_for	ATTAGGTTTATGCGGTCTGTT	21mer	*coxI*	*C. velia*	52
P772	Vb1_CoxI_for	TGTAGTGGCCCATTTTCATTT	21mer	*coxI*	*V. brassicaformis*	59
P773	Vb2_CoxI_for	AGGATTTCAATTAATGGTAAGAAGGA	26mer	*coxI*	*V. brassicaformis*	60
P774	Vb1_CoxI_rev	CCACCAAATTCTGAATGTATAGATAA	26mer	*coxI*	*V. brassicaformis*	58
P775	Vb2_CoxI_rev	CTGATATAACCAAGTATCTTTGTGTGG	27mer	*coxI*	*V. brassicaformis*	59
P801	Pols_CoxI_for2	AGGTTTTAATGTAATGCCCCTA	22mer	*coxI*	*P. olseni*	57
P845	Pols_CoxI_for3	TAGGCTTTAGGGATCAACCA	20mer	*coxI*	*P. olseni*	56
P803	Pols_CoxI_rev2	GGGGTATACCACCATATAAACC	22mer	*coxI*	*P. olseni*	60
P805	Pols_COB_for2	ACACCAATACATATAATACCAGAATGA	27mer	*cob*	*P. olseni*	61
P806	Pols_COB_rev1	AACATTATCATACCAGAATATCA	23mer	*cob*	*P. olseni*	54
P807	Pols_COB_rev2	ACTTGTACAATCAATTACTACCAT	24mer	*cob*	*P. olseni*	57

Table S3: Assembly of established primers utilized in tailing PCR experiments for the review of a possible linearity of the mitochondrial genome in the chromerid *C. velia*.

Tailing PCR experiments: Established primer pairs for the *C. velia* mitochondrion						
#	**Name**	**Sequence**	**Length**	**Gene**	**Organism**	**Tm [°C]**
P668	Cv_Mito_rev	CCTAAACCAACACCAAAATCT	21mer	*coxI*	*C. velia*	51
P669	Cv_Mito_for	TACAGCTATTTTATCTTATTT	21mer	*coxI*	*C. velia*	41
P734	TdT_Anker_dG	CTACTACTACTAGGCCACGCGTCGACTAGTACGGGGGGGGGGGGGGGG	48mer		universal	75
P735	TdT_Anker_dT	CTACTACTACTAGGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTT	48mer		universal	70
P736	TdT_Anker_uni	CTACTACTACTAGGCCACGCGTCGACTAGTAC	32mer		universal	67

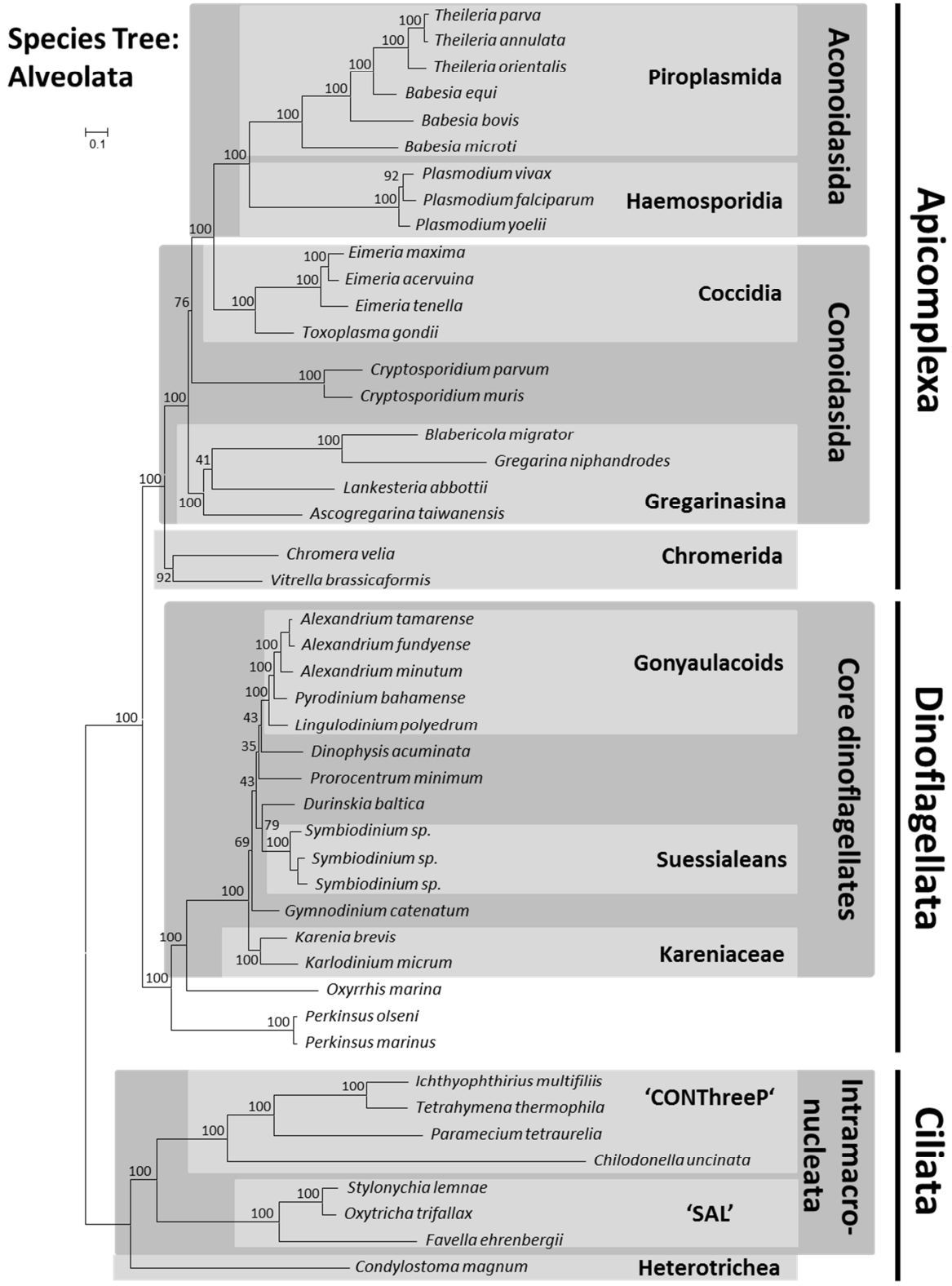


Figure S3: Phylogenomic tree incorporating 46 alveolate species. The analysis is calculated by the CAT-GTR method and is based on more than 150 concatenated orthologous nuclear genes with 30,065 aa positions. The phylum Ciliata forms the outgroup of this tree. The bootstrap values are according to the jackknife resampling method.

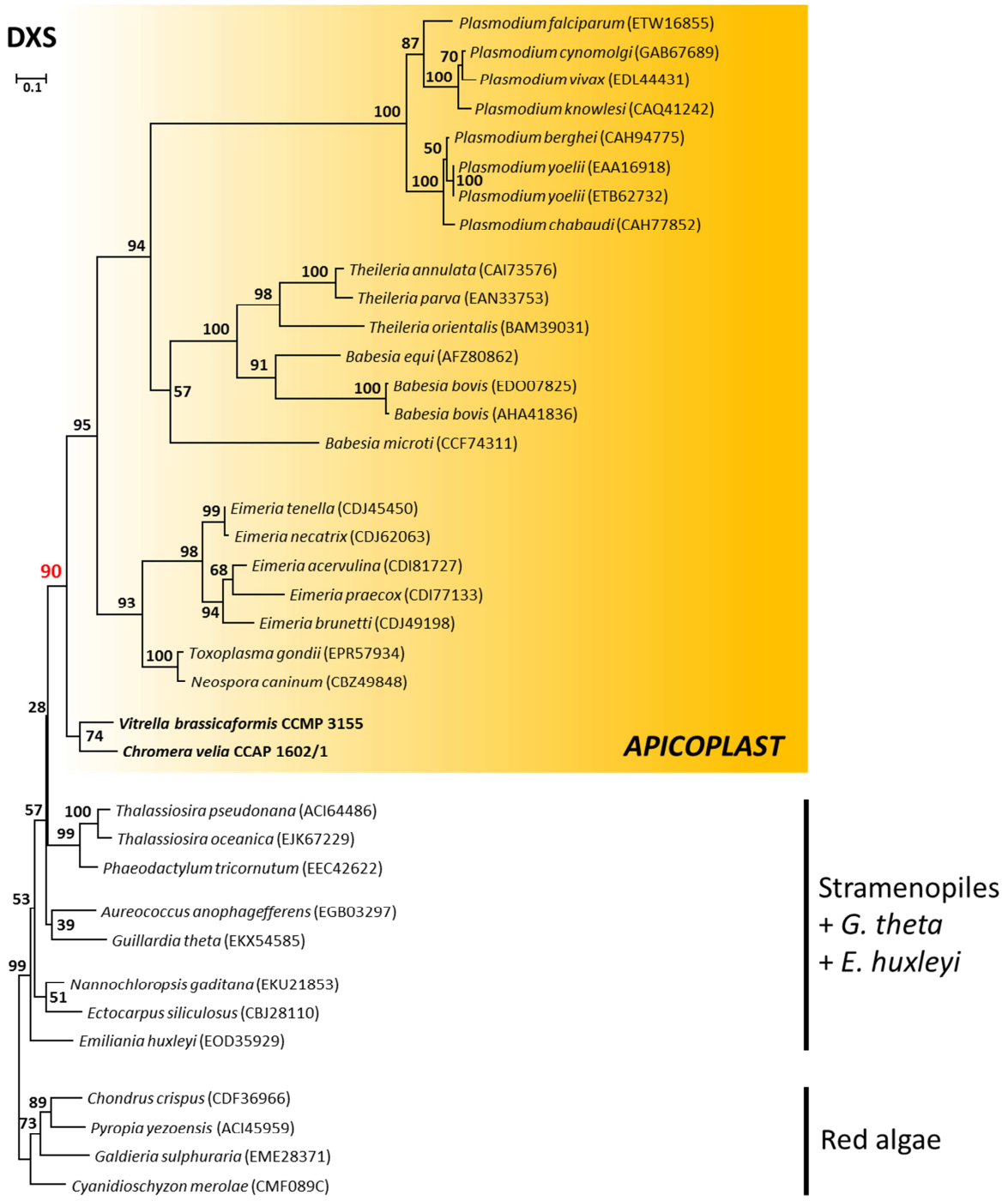


Figure S4: Extensive phylogenetic RAxML (LG+F+4Γ) tree based on 36 species and 366 aa positions. Sequences representing the apicoplast are highlighted by a yellow box. The bootstrap for the apicoplast monophyly is marked in red. Sequences of transcriptomes established in this study are shown in bold.

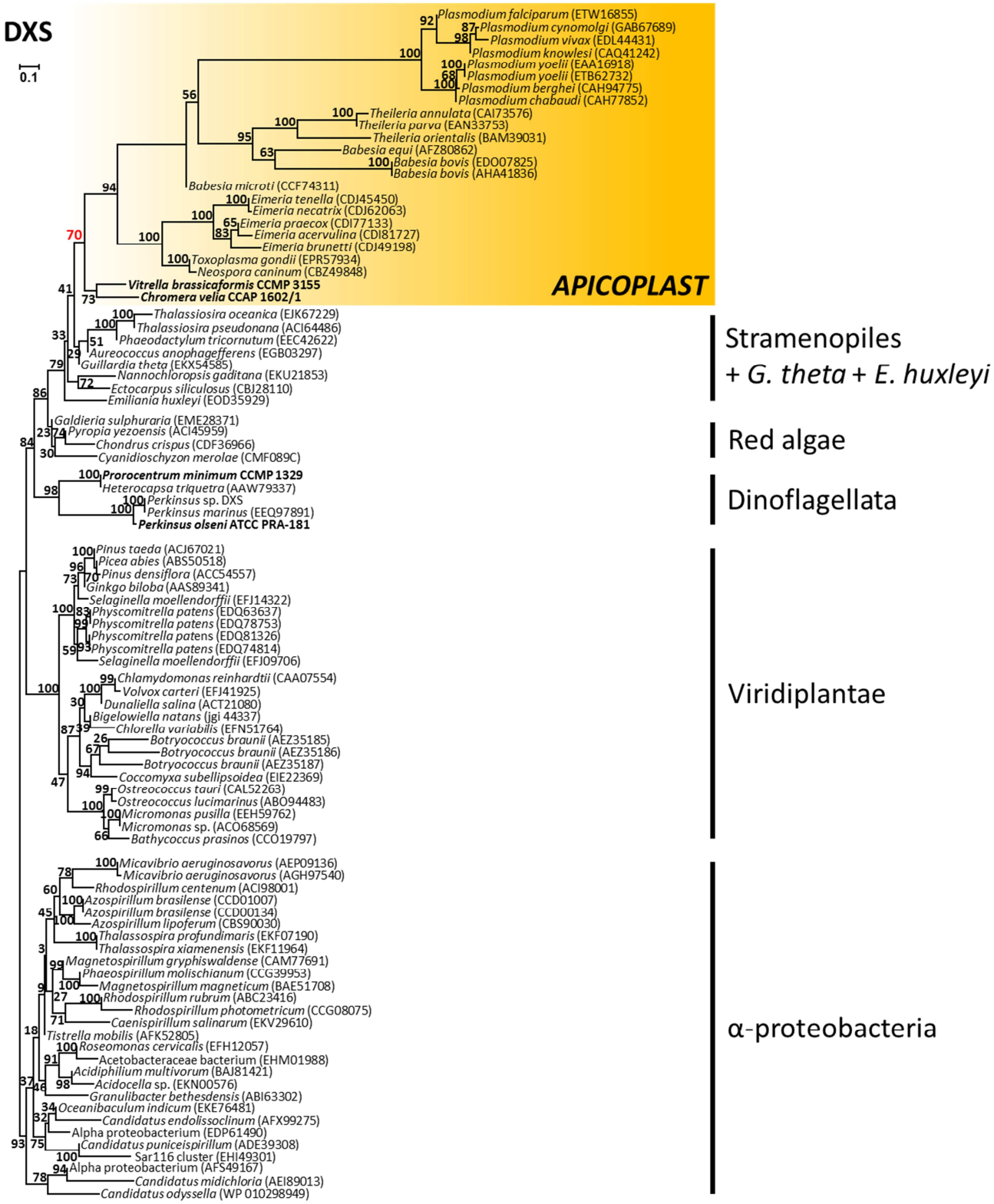


Figure S5: Phylogenetic RAxML (LG+F+4Γ) tree based on 93 species and 450 aa positions. Sequences representing the apicoplast are highlighted by a yellow box. The bootstrap for the apicoplast monophyly is marked in red. Sequences of transcriptomes established in this study are shown in bold.

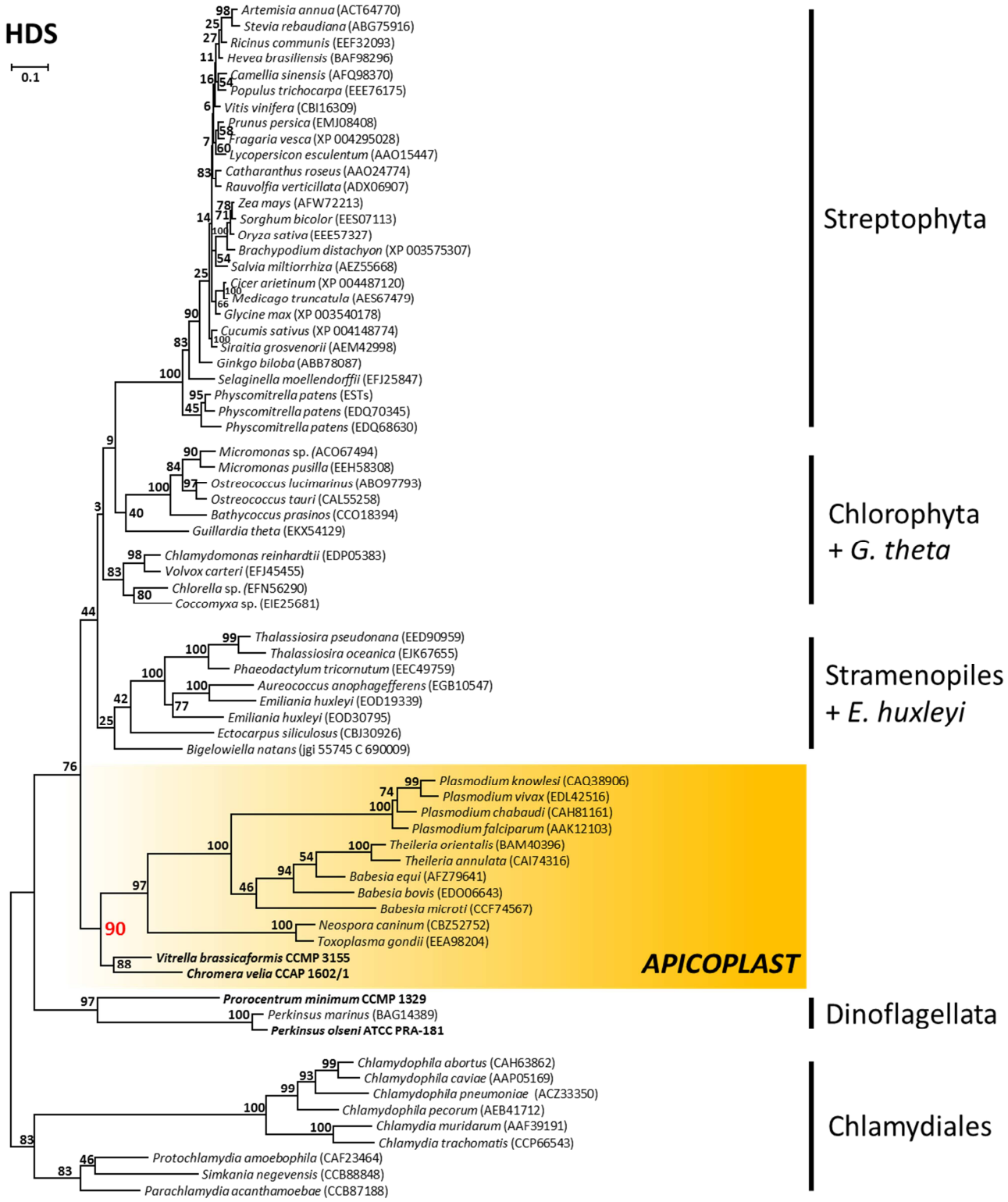


Figure S6: Extensive phylogenetic RAxML (LG+F+4Γ) tree based on 70 species and 462 aa positions. Sequences representing the apicoplast are highlighted in a yellow box. The bootstrap for the monophyly of the apicoplast is marked in red. Transcriptomic sequences established in this study are shown in bold.

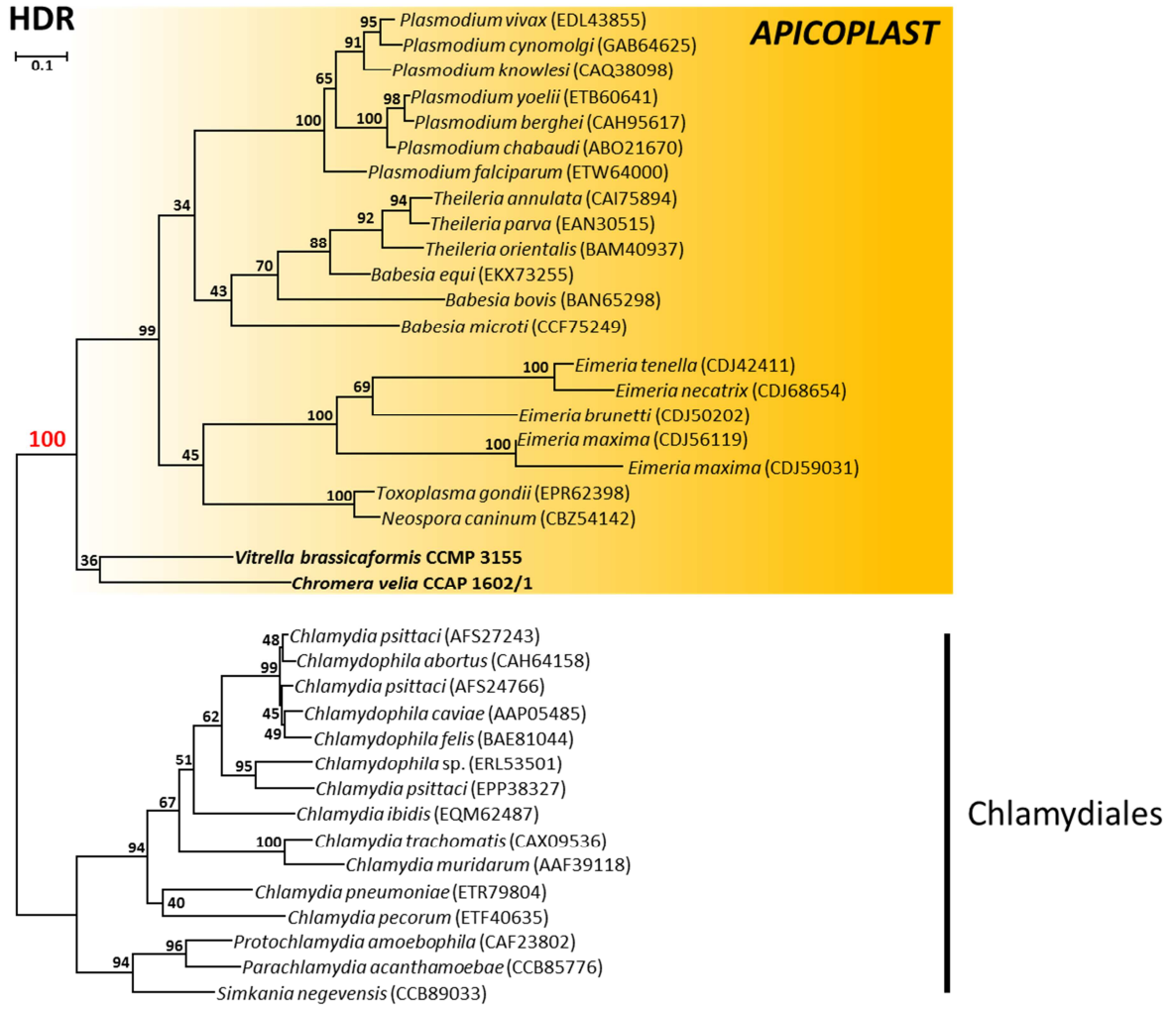


Figure S7: Extensive phylogenetic RAxML (LG+F+4Γ) tree based on 37 species and 241 aa positions. Sequences representing the apicoplast are highlighted by a yellow box. The bootstrap value for the monophyly of the apicoplast is marked in red. Transciptomic sequences established in this study are shown in bold.

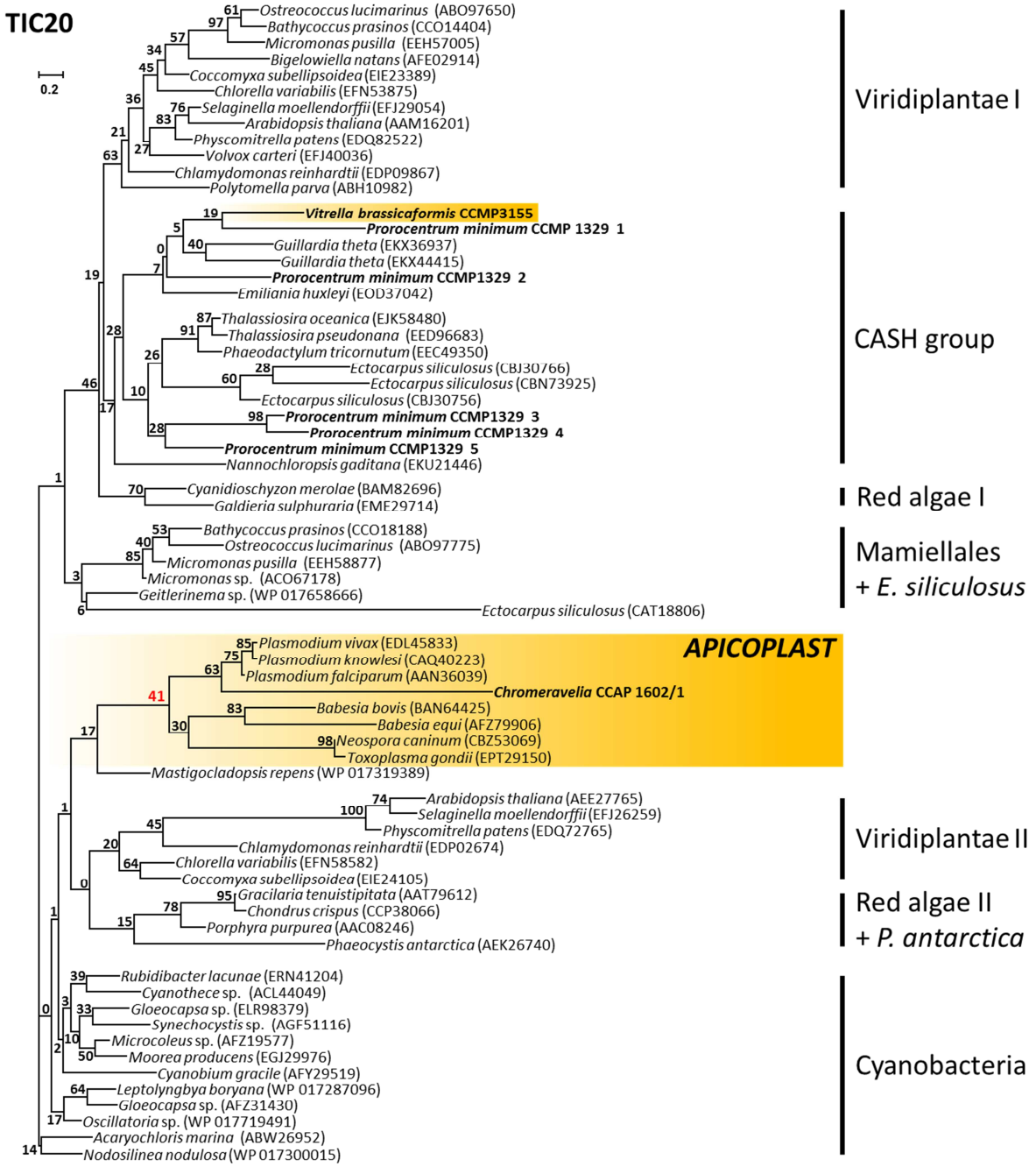


Figure S8: Phylogenetic RAxML (LG+F+4Γ) tree based on 67 species and 82 variable aa positions. Apicomplexan sequences are highlighted by a yellow box. The bootstrap value for the apicoplast group (*V. brassicaformis* excluded) is marked in red. Alveolate transcriptomic sequences established in this study are shown in bold.

Table S4: Seven sequences of the transcriptomic dataset of *P. olseni* annotated in a photosynthetic context by KAAS. These results were counterchecked by manual blastx analyses. The first hit of each blastx result is shown accompanied by its respective E-value. The last column represents the annotation result for each sequence calculated by the Blast2GO pipeline.

KAAS annotation	Gene	*P. olseni*	blastx result (1st hit)	E-value	Blast2GO annotation
Photosystem II cytochrome b559 subunit alpha	psbE	extcontig11101	protein phosphatase 2C, putative (*P. marinus*, XP_002777897)	$3e^{-165}$	protein phosphatase [EC 3.1.3.16; 3.1; 3.1.3.41]
Photosystem II 13kDa protein	psb28	contig1006	hypothetical protein (*P. marinus*, Pmar_PMAR001448)	0.033	hypothetical protein Pmar_PMAR 005291
Photosystem I P700 chlorophyll a apoprotein A1	psaA	extcontig5264	hypothetical protein (*P. marinus*, Pmar_PMAR020715)	0.0	hypothetical protein Pmar_PMAR 020715
Photosystem I subunit	psaO	extcontig6299	conserved hypothetical protein (*P. marinus*, XP_002782381)	$3e^{-115}$	signal peptide containing protein
Phycobilisome rod-core linker protein	cpcG	contig1614	hypothetical protein (*Aphanomyces astaci*, H257_17072)	$1e^{-62}$	isovaleryl-dehydrogenase (ivd)
Light-harvesting complex II chlorophyll a/b binding protein 5	LHCB5	contig1268	conserved hypothetical protein (*P. marinus*, XP_002786249)	$7e^{-159}$	hypothetical protein Pmar_PMAR 018037
Light-harvesting complex II chlorophyll a/b binding protein 7	LHCB7	contig2860	PREDICTED: ryanodine receptor 1-like (*Esox lucius*, XP_012991024)	0.007	--- NA ---

Table S5: List of essential plastid pathway genes (Lim and McFadden 2010) in the parasitic dinoflagellate *P. olseni* and results of the Venn diagram analysis. Shown is the presence of these genes in the respective genomic dataset of *P. olseni* and their coverage values as well as the presence of introns (check mark). Absence of genes and introns is marked by a dash. The results are due to local BLAST analyses with appropriate query sequences (e.g. *T. thermophila*, *T. gondii*, *A. thaliana*, *C. velia* and *Hematodinium* sp.) in the genomic A3-dataset of *P. olseni*. p = plastid localization , m = mitochondrial localization, c = cytosolic localization (predicted by TargetP and iPSORT). Abbreviations: MEP = Non-mevalonate pathway; SUF = iron-sulphur cluster assembly; FAS = Fatty acid biogenesis; TP = Tetrapyrrole.

Pathway/Function	Gene	*P. olseni* genomic Node	Coverage	Introns
MEP	*dxs*^{p}	1564	117x	✓
	dxr^{p}	2374	97x	✓
	mct^{p}	2396/4684	202x/179x	✓
	cmk^{p}	2264	126x	✓
	mecps^{p}	2026	127x	✓
	hds^{p}	600/3845	110x/127x	✓
	hdr^{p}	4197	118x	✓
SUF system	*sufA*^{p}	1283/3912	96x/110x	✓
	sufB^{p}	429	217x	✓
	sufC^{p}	2758	105x	✓
	sufD^{p}	4929	107x	✓
	sufE^{p}	2451	116x	✓
	sufS^{p}	4929	107x	✓
FAS II	*acp*	-	-	-
	fabB	-	-	-
	fabF	-	-	-
	fabD	-	-	-
	fabH	-	-	-
	fabI	-	-	-
	fabG^{m}	4766	124x	✓
	fabZ	-	-	-
FAS I	Multienzyme^{c}	286	101x	✓
TP biosynthesis	ALAS	1097	134x	-
		10783	104x	-
	hemB^{p}	672	137x	✓
	hemC^{p}	6870	114x	✓
	hemD	-	-	-
	hemE^{p}	752	90x	✓
	hemF	-	-	-
	hemY/G^{m}	2891	87x	✓
	hemH	-	-	-
Chaperone	*clpC*^{c}	1704	106x	✓

Table S6: List of plastid-encoded genes in the red alga *Chondrus crispus* (***Cc***) and the presence/absence of homologs in the genome of the apicomplexan parasite *P. falciparum* (***Pf***), as well as the presence/absence of paralogs in the genomes of the aplastidial alveolate species *Cryptosporidium parvum* (***Cp***) and *Tetrahymena thermophila* (***Tt***). Presence is marked by the appropriate Accession number of the NCBI database; absence is marked by a dash. In the case of homologs in *P. falciparum*, but no paralogs in the aplastidial species (yellow), local blast analyses in the genome of *P. olseni* (***Po***) were accomplished. The presence of homologs in *P. olseni* is specified by the number of the correspondent node in the genomic assembly. Furthermore, the coverage value to each genomic node is listed, as well as the presence of introns highlighted by a check mark.

Gene	*Cc* plastome	*Pf* genome	*Cp* genome	*Tt* genome	*Po* genome	Coverage [x]	Introns
Biosynthesis							
accA	YP_007627470.1	-	-	-			
accB	YP_007627396.1	-	-	-			
accD	YP_007627299.1	-	-	-			
acpP	YP_007627399.1	XP_001349595.1	-	XP_001470926.1			
acsF	YP_007627401.1	-	-	-			
argB	YP_007627351.1	-	-	-			
carA	YP_007627302.1	XP_001349809.1	-	-	614	108	✓
ccs1	YP_007627316.1	-	-	-			
chlI	YP_007627492.1	-	-	-			
dsbD	YP_007627315.1	-	-	-			
fabH	YP_007627297.1	XP_001349620.1	-	-	-	-	-
gltB	YP_007627474.1	XP_001348508.1	-	XP_001026023.1			
hisS	YP_007627331.1	-	XP_627292.1	XP_001014072.3			
ilvB	YP_007627349.1	-	-	-			
ilvH	YP_007627456.1	-	-	-			
moeB	YP_007627343.1	-	XP_626339.1	XP_001032732.3			
preA	YP_007627410.1	XP_001349541.1	XP_628536.1	XP_001032054.2			
rbcL (red)	YP_007627452.1	-	-	-			
rbcS(red)	YP_007627453.1	-	-	-			
syfB	YP_007627333.1	-	-	XP_001020513.3			
thiG	YP_007627318.1	-	-	-			
thiS	YP_007627335.1	-	-	-			

Gene	*Cc* plastome	*Pf* genome	*Cp* genome	*Tt* genome	*Po* genome	Coverage [x]	Introns
tilS	YP_007627482.1	-	-	-			
trpA	YP_007627481.1	-	-	-			
trpG	YP_007627317.1	XP_001352095.1	-	-	3487	134	✓
trxA	YP_007627451.1	XP_001348719.1	XP_62576.1	XP_001022530.1			
upp	YP_007627325.1	-	-	-			
Transport							
cemA	YP_007627444.1	-	-	-			
secA	YP_007627480.1	-	-	XP_001007094.1			
secG	YP_007627404.1	-	-	-			
secY	YP_007627371.1	-	-	-			
sufA (ycf57)	YP_007627460.1	XP_001351784.1	-	XP_001010095.2			
sufB	YP_007627437.1	CAA64569.1	-	-	429	117	✓
sufC	YP_007627436.1	SBT78931.1	-	-	2758	105	✓
tatC	YP_007627414.1	-	-	-			
Cell processes							
bas1	YP_007627406.1	-	XP_001388054.1	XP_001031885.2			
clpC	YP_007627346.1	SBT77047.1	-	-	1704	106	✓
dnaB	YP_007627345.1	-	-	-			
dnaK	YP_007627389.1	SBT77217.1	XP_626895.1	XP_001030821.2			
ftsH	YP_007627352.1	XP_001350791.1	XP_628150.1	XP_001470868.2			
groEL	YP_007627330.1	XP_001347438.1	XP_627821.1	XP_001017291.2			
nblA	YP_007627465.1	-	-	-			
psbA	YP_007627313.1	-	-	-			
rbcR	YP_007627306.1	-	-	-			
Energy metabolism							
atpA	YP_007627435.1	SBT75623.1	XP_001388301.1	XP_001027692.2			
atpB	YP_007627419.1	XP_001350751.1	XP_001388200.1	XP_001032631.2			
atpD	YP_007627434.1	-	-	-			
atpE	YP_007627418.1	-	-	-			

Gene	*Cc* plastome	*Pf* genome	*Cp* genome	*Tt* genome	*Po* genome	Coverage [x]	Introns
atpF	YP_007627433.1	-	-	-			
atpG	YP_007627432.1	-	-	-			
atpH	YP_007627431.1	-	-	-			
atpI	YP_007627430.1	-	-	-			
cbbX	YP_007627454.1	-	-	-			
ftrB	YP_007627485.1	-	-	-			
pdhA	YP_007627411.1	XP_001347927.1	-	XP_001017076.2			
pdhB	YP_007627412.1	XP_001348615.2	-	XP_001016910.2			
petA	YP_007627413.1	-	-	-			
petB	YP_007627338.1	-	-	-			
petD	YP_007627339.1	-	-	-			
petF	YP_007627359.1	XP_001349917.1	-	-	1424	102	✓
petG	YP_007627392.1	-	-	-			
petL	YP_007627447.1	-	-	-			
petN	YP_007627402.1	-	-	-			
pgmA	YP_007627478.1	-	-	-			

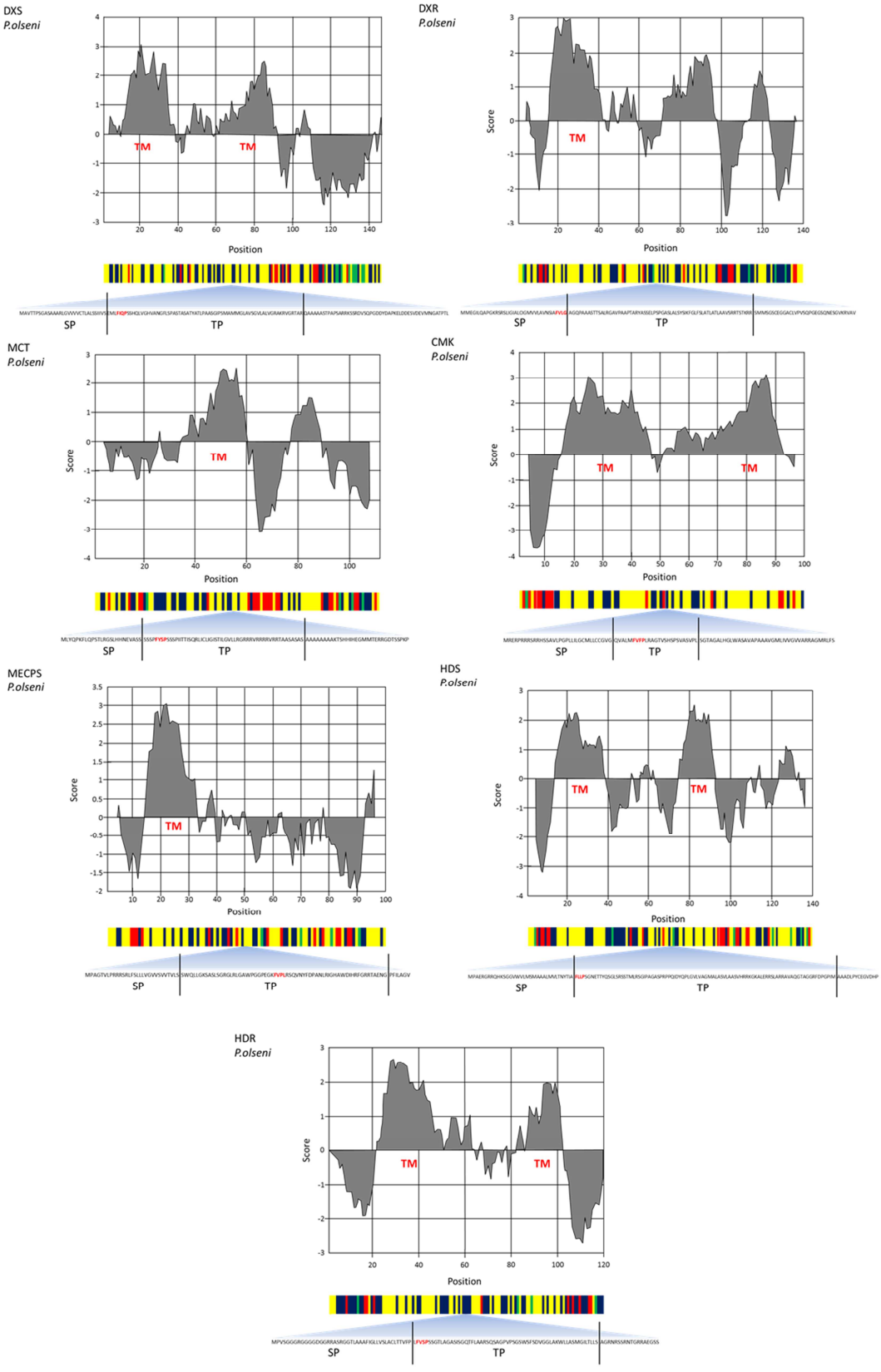


Figure S9: Inferred plastidial bipartite signal sequences for the first >100aa residues of the seven MEP proteins DXS, DXR, MCT, CMK, MECPS, HDS and HDR. Shown are the protein sequences with their calculated Kyte-Doolittle hydropathy plots. The amino acids are classified into groups (yellow: hydrophobic, non-polar; blue: polar, hydrophilic; red: basic, positively charged; green: acidic, negatively charged). The transmembrane domains in the Kyte-Doolittle plots as well as the conserved 'FVAP' motif are highlighted in red. Abbreviations: SP = signal peptide; TP = transit peptide; TM = transmembrane domain.

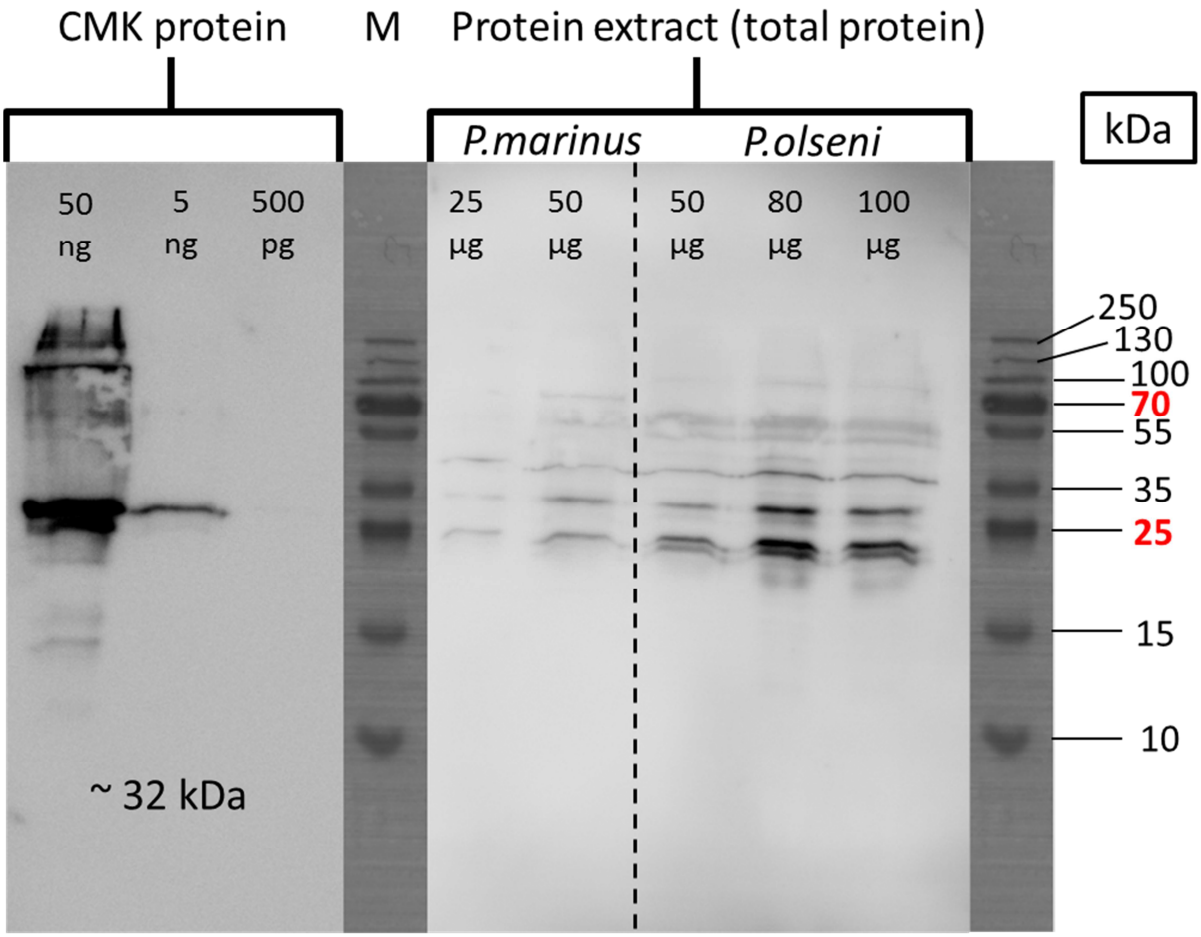


Figure S10: Western Blot analysis for the binding specifity of the antibody RB3153 ('CMK_1') against the purified MEP-protein CMK and against the total protein extract from *P. marinus* and *P. olseni* cell cultures.

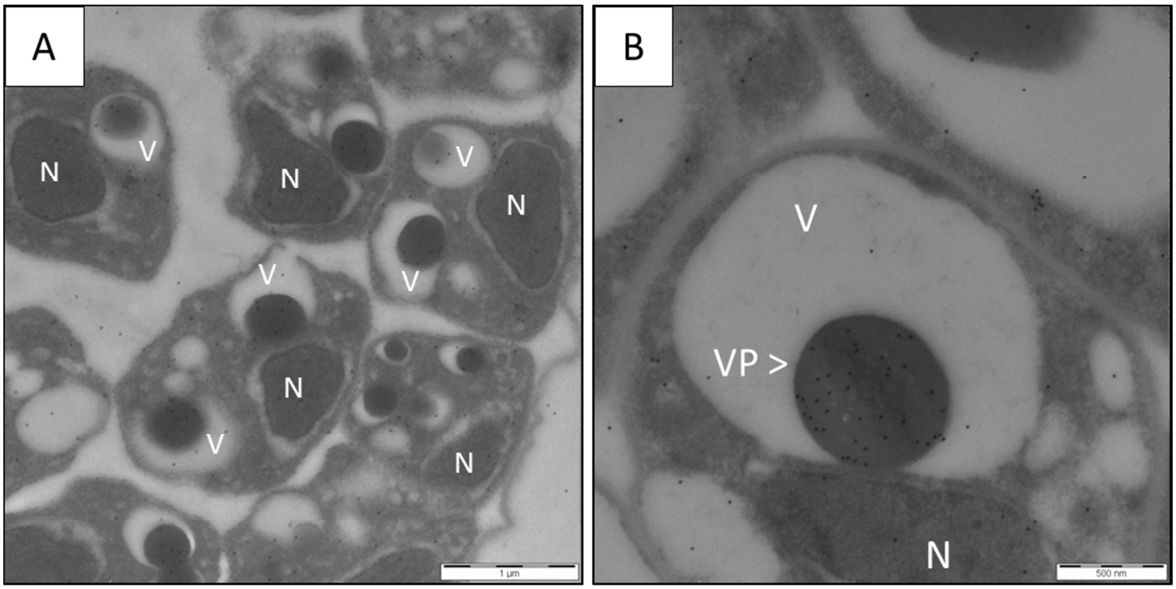
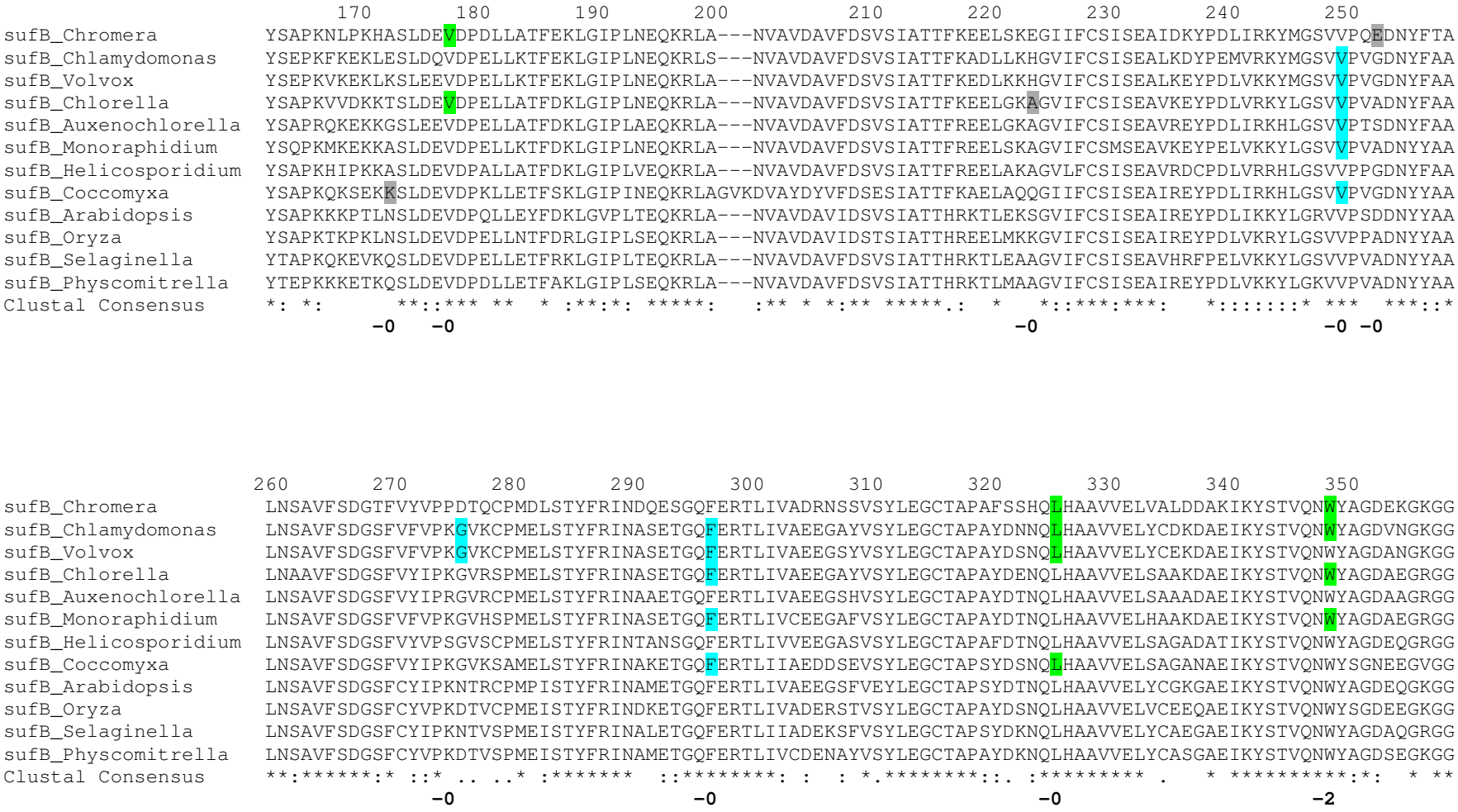


Figure S11: Electron micrographs of zoospore-enriched cells of *P. olseni*. The samples were incubated with the CMK_1 antibody which binding specifity was detected via protein A linked to 15 nm colloidal gold. A) Round-shaped cells (~1.5-2.5 µm) with large nuclei (N) and numerous vacuoles (V) containing electron dense structures. Unspecific binding of the CMK_1 antibody. B) Accumulations of the CMK_1 antibody in such an electron dense structure (possibly a vacuoplast, 'VP').

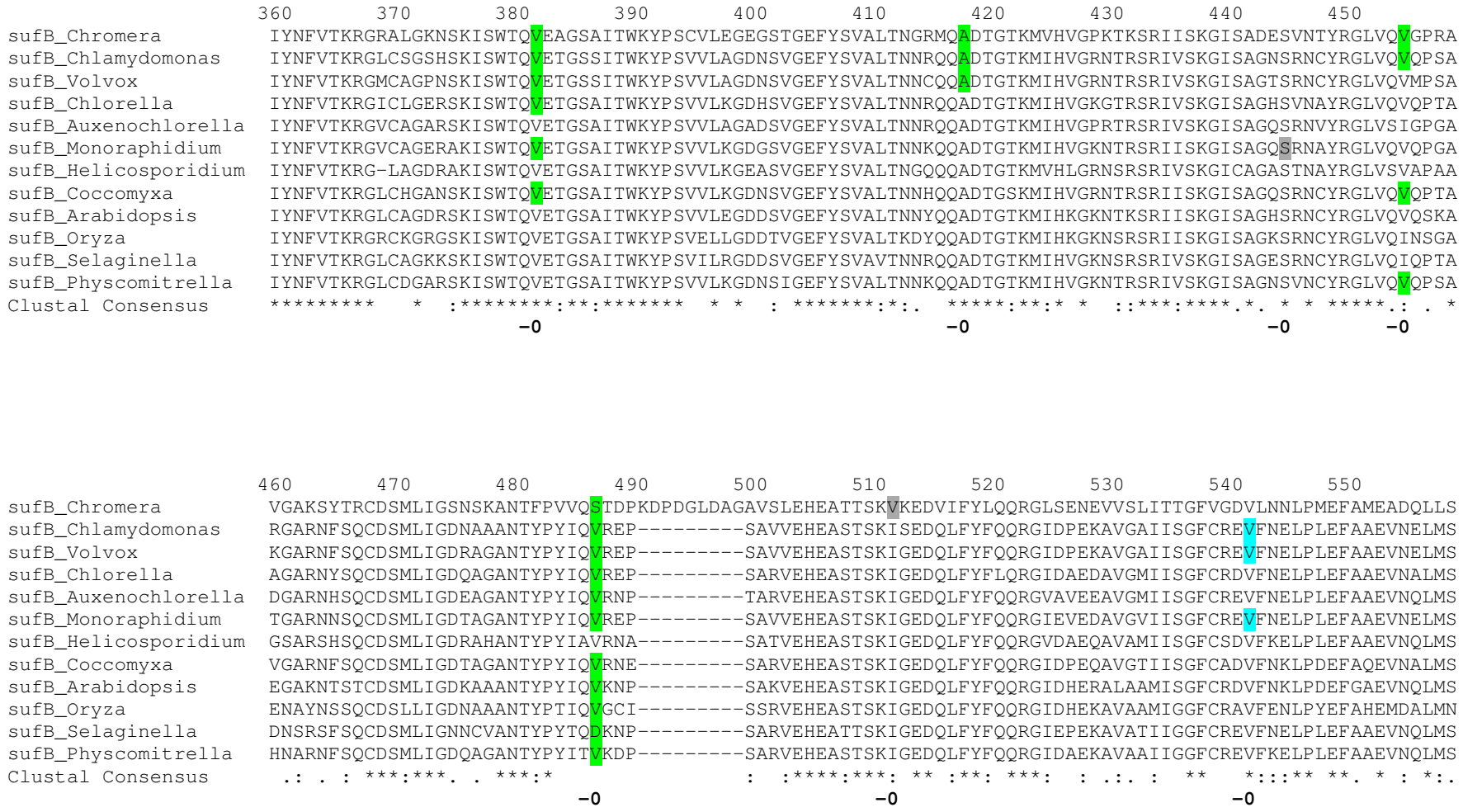
SufB-Alignment: Intron Distribution and Conservation

```
                                                                      10        20        30        40        50        60
sufB_Chromera                                                MLFLGLASGFALLTGSEAFRLVPSRSSEGGREGRLFKRGVSRAGDRDRHDDVPSRLFEAPP
sufB_Chlamydomonas                                                                  MSASMLPCGTRTCSGASTSTRTQRAAAPGVPIAARSLR
sufB_Volvox                                                                        MVSLLPSRTCSGASTSECGRRNAATVSCGIRTSYRSQPH
sufB_Chlorella
sufB_Auxenochlorella
sufB_Monoraphidium
sufB_Helicosporidium
sufB_Coccomyxa
sufB_Arabidopsis                                             MASLLANGISSFSPQPTSDSSKSPKGFHPKPESLKFPSPKSLNPTRPIFKLRADVGIDSRP
sufB_Oryza                                                                MAAAAASSTPLFSPCCAAATAKLGAACPSSYGSRRRPCTRRGLSVVA
sufB_Selaginella                                                                    MPQPAIGFLGGRLYQVRKSLEIPRKSRRFSSIRAEV
sufB_Physcomitrella   MAMVVDIGVVATPAAAVVCKCKHGGASLPTSSSACVVGMGGGHALPVCRGLRRLSVAAKRPEPVGFQARSSRSVGIVAEVQEAASPVFESPVKGSSDDTI
Clustal Consensus

                              70        80        90       100       110       120       130       140       150       160
sufB_Chromera         SSTSTLEIEPRAIPKEVSVEDKFFSAANRAYQYGFTSKIESTSLPPGLDEEKLRQLSAIKEEPQWMLDWRLKAYMRWKKMKEPQWAHVYYPPVNYDTMSY
sufB_Chlamydomonas    RGPTTCRTATVEEAQISDEKRDIQKILDRPYKYGFKTIIESDTFPKGLNEDVVRAISAKKGEPEWMLEFRLKAFRKWLTMEEPKWSDNAYPEIDYQDVSY
sufB_Volvox           QRAMSCRTAMIEDTQLSDEKRDIQKILDRPYKYGFKTIIESDTFPKGLNEDVVRAISAKKGEPEWMLEFRLKAYRKWLTMEEPKWSDNAYPEIDYQDVSY
sufB_Chlorella                                 MLAKPYKYGFKTIIETEQFPKGLSEDVVRAISAKKEEPEWMLDFRLKAYRKWLTMEEPNWSDNRYPRIDYQDLSY
sufB_Auxenochlorella                           MLSKPYKYGFKTFIESDVFPKGLDEDVVRAISLKKEEPEWLLEFRLKAWRRWLAMEEPTWSDNDYPAIDYQAYSY
sufB_Monoraphidium                             MLNRPYKYGFQTFIESDTFPKGLDEDVVRAISAKKGEPDWMLEFRLRAFRKWLTLEEPRWSDNQHPVIDFQDLSY
sufB_Helicosporidium                            MSKPYKYGFQTEIESDVFPKGLSEDVVRAISAKKEEPEWMLDFRLKAYRRWLTMEEPCWSDVRFPPIDYQSFTY
sufB_Coccomyxa                             MLSGPCREYTPGFTTLIESETFAKGLDETVVRAISAKKREPEWMLDFRLKAYRKWLTMAEPAWSDNRYPQINFQDLSY
sufB_Arabidopsis      IGASESSSSGTSTVSSTDKLQQYFQNLDYDKKYGFVEDIDSFTIPKGLSEETIRLISKLKEEPDWMLEFRFKAYAKFLKLEEPKWSDNRYPSINFQDMCY
sufB_Oryza            VQTGPQKPSPSSSSQAGTESETLQNLLKREYKYGFVSDFESFSIPKGLSEATVRRISELKAEPAWMLDFRLAAYRRFLTMVQPTWSDNVYEPVDLQSICY
sufB_Selaginella      PPPEVLEIKPSSSGTGSSSDETIRQFLDRDYKWGFVSNVASESIPKGLDESTVRLISSKKKEPDWLLQFRLNAYRQWLKMEEPSWSDNTYPKIDFQDFCY
sufB_Physcomitrella   QQFLKRDYKYGFVSDIGSSDDTIQQFLKRDYKYGFVSDIESVSIPKGLSEETVRLISAKKNEPEWMLNFRLNAFKQWQKMKEPTWSDNKYPQIDFQDVCY
Clustal Consensus                                      **   . :  :. **.*  :* :*  * ** *:*::*: *: ::  : :* *:.  .  :: :   *
                                                                                    -0    -2   -2            -2
```

SufB-Alignment: Intron Distribution and Conservation



SufB-Alignment: Intron Distribution and Conservation



SufB-Alignment: Intron Distribution and Conservation

	Organism	Strain	Introns	(-0	-1	-2)	Protein(Σaa)
	560						
sufB_Chromera	VKLSNPVG*Chromera velia*	CCAP 1602/1	10	8	0	2	566
sufB_Chlamydomonas	LKLEGTVG*Chlamydomonas reinhardtii*	CC-503 cw92 mt+	10	9	0	1	534
sufB_Volvox	LKLEGTVG*Volvox carteri*	f. nagariensis/Eve	8	8	0	0	535
sufB_Chlorella	LKLEGSVG*Chlorella variabilis*	NC64A	8	5	0	3	471
sufB_Auxenochlorella	LKLEGSVG*Auxenochlorella protothecoides*	0710	3	2	0	1	471
sufB_Monoraphidium	LKLEGSVG*Monoraphidium neglectum*	SAG 48.87	7	6	0	1	471
sufB_Helicosporidium	LKLEGSVG*Helicosporidium sp.*	ATCC 50920	1	0	0	1	469
sufB_Coccomyxa	LKLEGSVG*Coccomyxa subellipsoidea*	C-169	9	8	0	1	477
sufB_Arabidopsis	IKLEGSVG*Arabidopsis thaliana*		1	1	0	0	557
sufB_Oryza	LKLEGSVG*Oryza sativa*	Os01g0830000	1	1	0	0	544
sufB_Selaginella	LKLENSVG*Selaginella moellendorffii*		1	1	0	0	532
sufB_Physcomitrella	LKLEGSVG*Physcomitrella patens*	"Gransden 2004"	2	2	0	0	596
Clustal Consensus	:**...**						

Intron legend:
- individual introns
- shared green algae
- shared Chromera/green algae

Clustal legend:
- `*` = single, fully conserved residue
- `:` = conservation between groups of strongly similar properties (>0.5 Gonnet PAM250 matrix)
- `.` = conservation between groups of strongly similar properties (≤0.5 Gonnet PAM250 matrix)

Figure S12: Protein alignment (ClustalW) of the plastidial marker SufB for *C. velia* and 11 streptophyte sequences. Coloured sites demonstrate the intron distribution within the SufB sequence of the respective organism as well as shared introns between the species. grey = introns only found in one species; blue = shared introns in green algae; green = shared introns between *C. velia* and green algae. Additionally information about the amount of introns and intron phase (-0, -1, -2) is recorded. Putative signal sequences at the N-terminus of the alignment are faded.

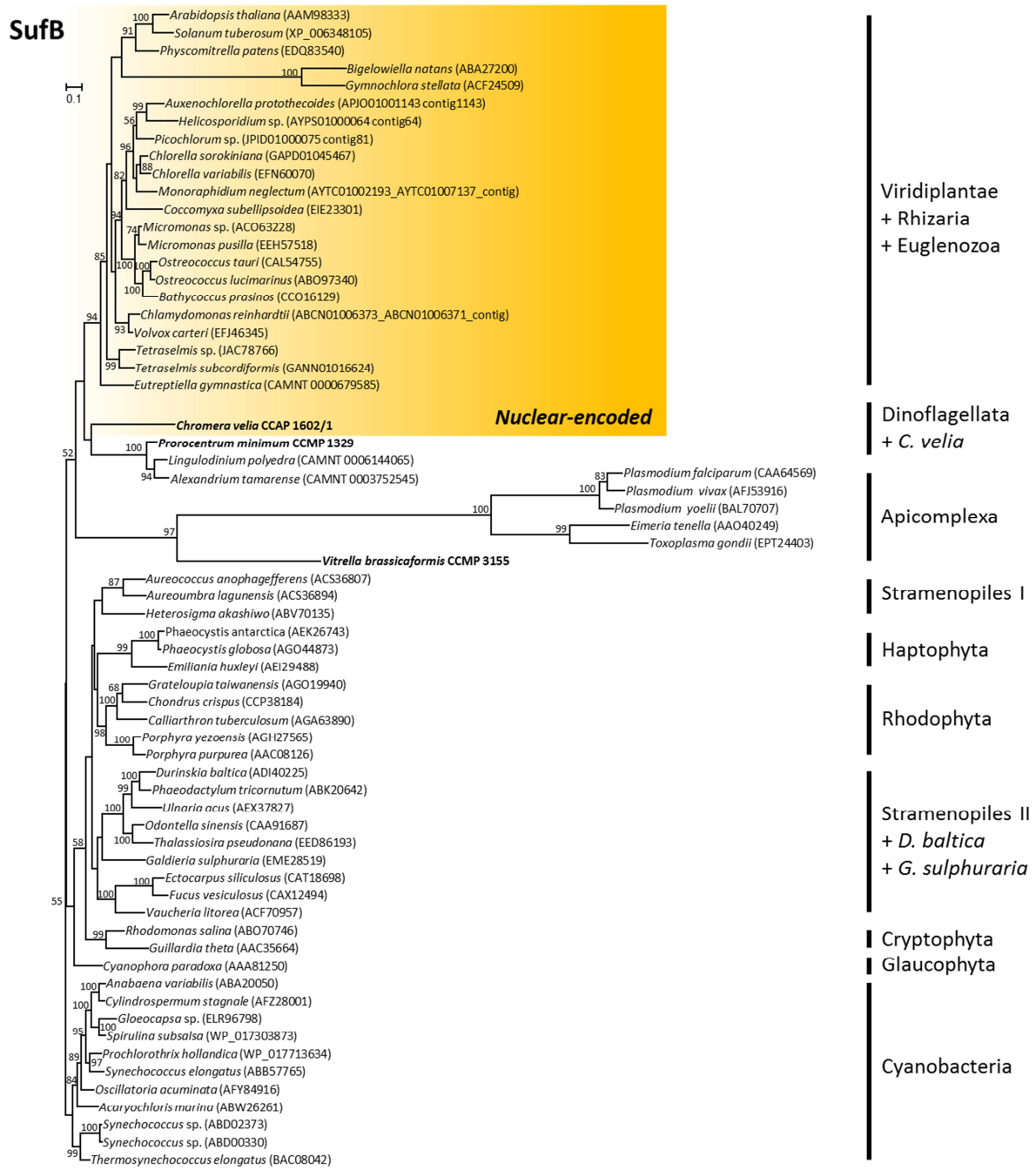
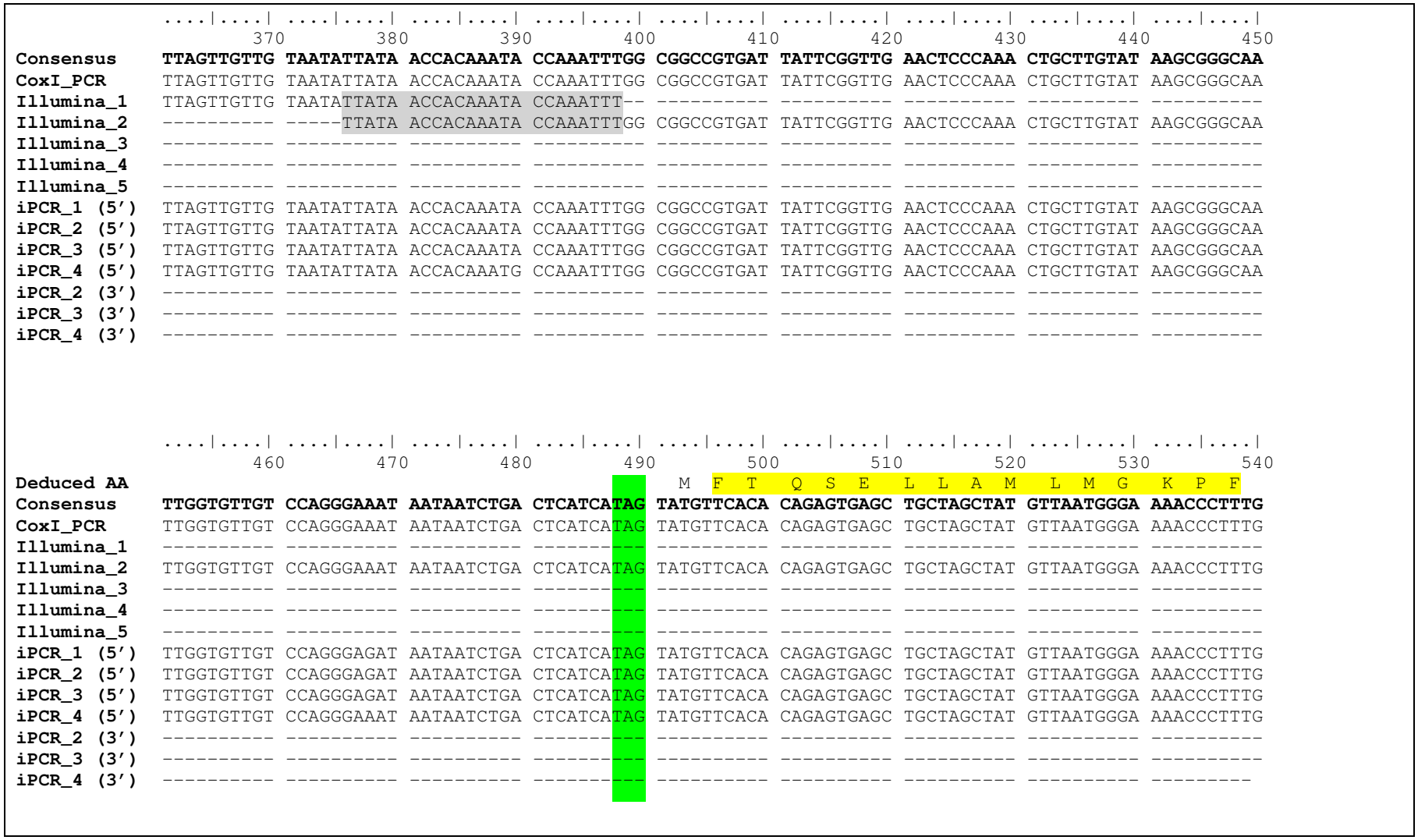


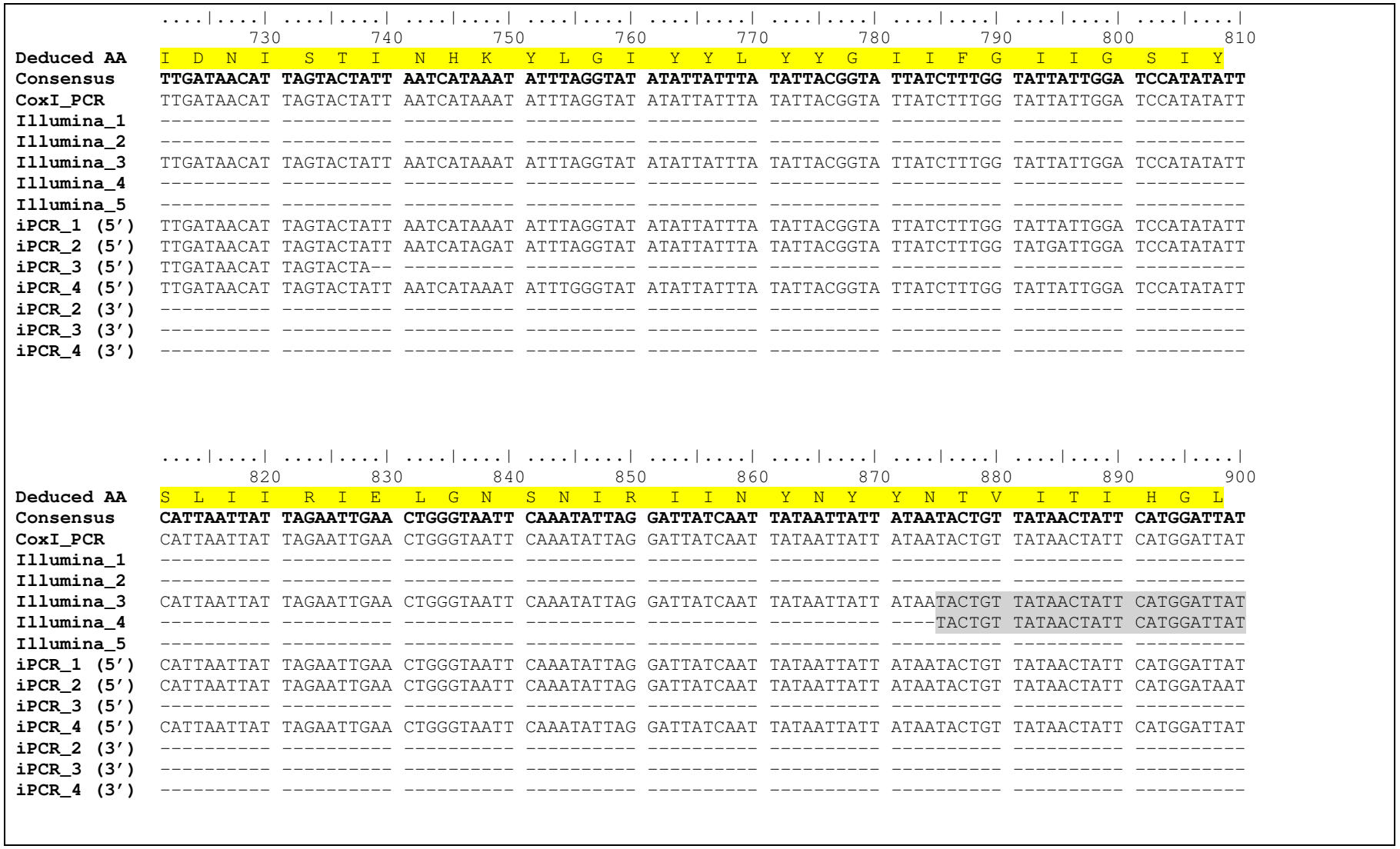
Figure S13: Phylogenetic tree based on a CATfix model (Le et al. 2008) with 20 pre-defined stationary probability profiles in PhyloBayes. The tree incorporates 66 eukaryotic species and 442 aa positions. Nuclear-encoded SufB sequences are highlighted in a yellow box. Sequences established in this study are marked in bold.

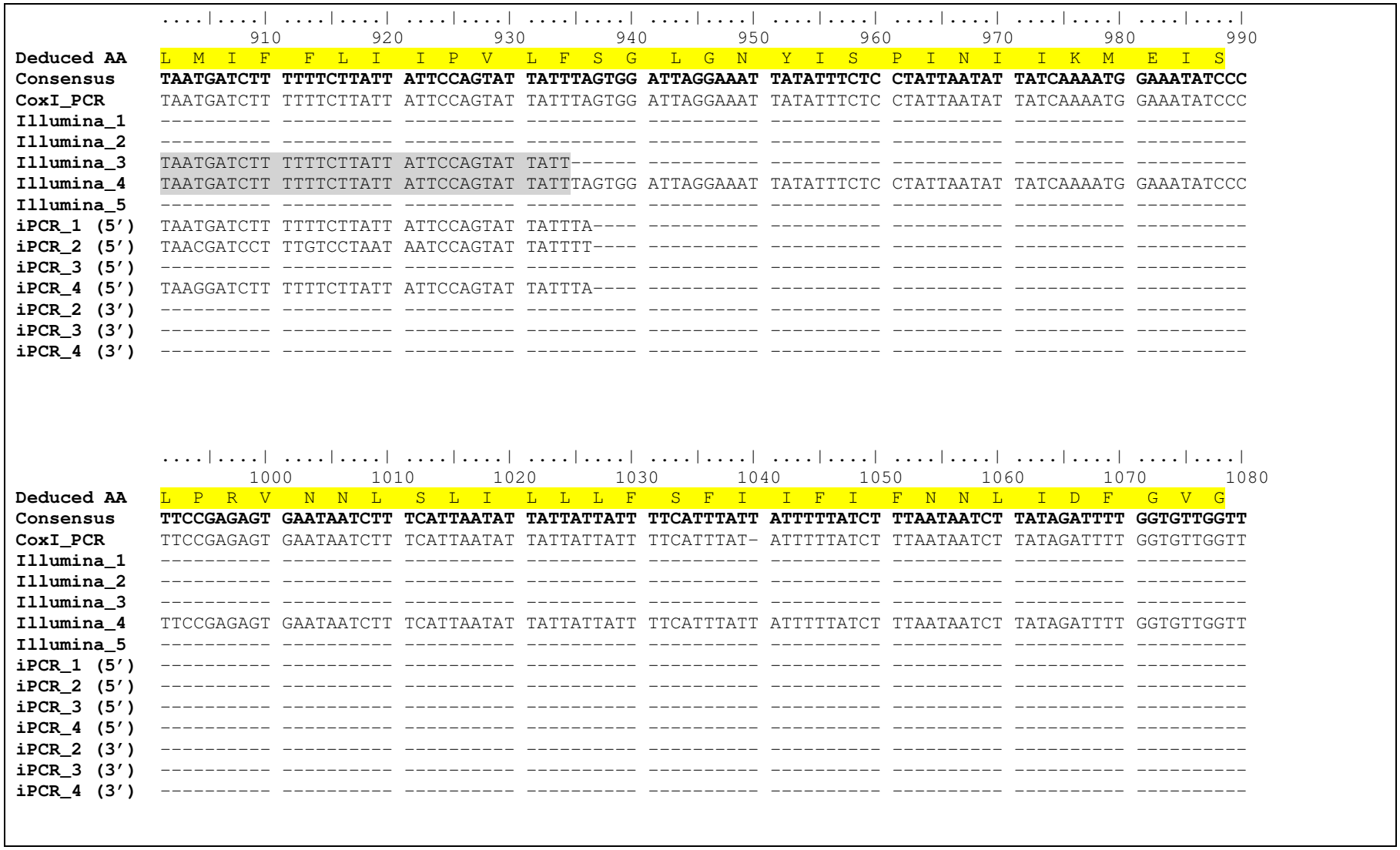
Mitochondrial genome reconstruction *C. velia* CCAP1602/1

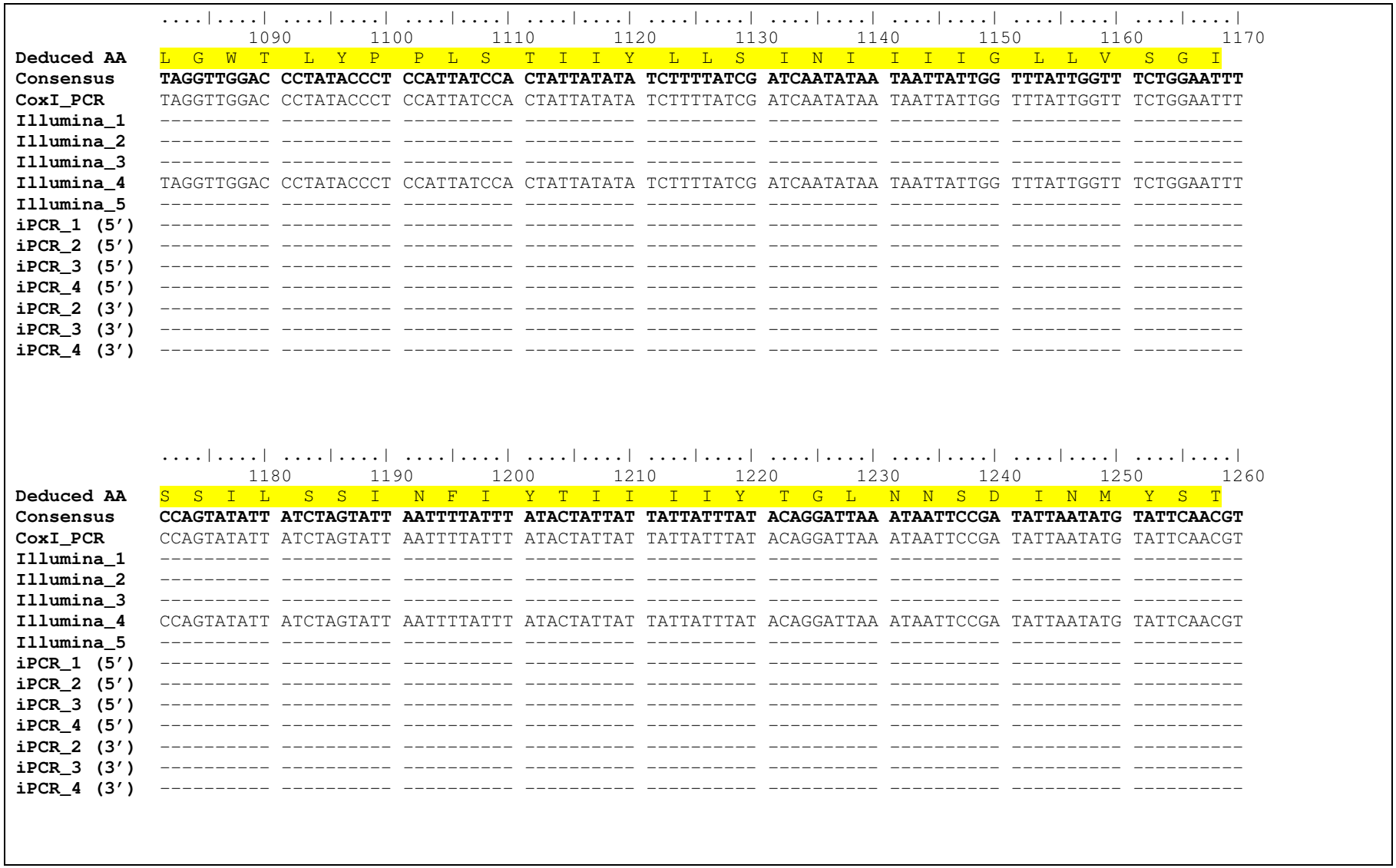

190 200 210 220 230 240 250 260 270
Consensus ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- TAGTTCAGAT
CoxI_PCR ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_1 ---------- ---------- -------TGT GTGGGTAAGG TTGGACAAGG ACTCCTGCAT TATATTGTGT GCCTCATCGC TAGTTCAGAT
Illumina_2 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_3 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_4 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_5 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_1 (5′) GTGTTGGATA AATATTAGAG TTGAACACAC ACAGATAACA CACACAGATA ATATTAATAG AACGGTTATA CGTGGGTTGG TAGTTCAGAT
iPCR_2 (5′) GTGTTGGATA AATATTAGAG TTGAACACAC ACAGATAACA CACACAGATA ATATTAATAG AACGGTTATA CGTGGGTTGG TAGTTCAGAT
iPCR_3 (5′) GTGTTGGATA AATATTAGAG TTGAACACAC ACAGATAACA CACACAGATA ATATTAATAG AACGGTTATA CGTGGGTTGG TAGTTCAGAT
iPCR_4 (5′) CATCCCTCTA TAGCTATATG TACTGTGTGT GTGGGTAAGG TTGGACAAGG ACTCCTGCAT TATATTGTGT GCCTCATCGC TAGTTCAGAT
iPCR_2 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_3 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_4 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
280 290 300 310 320 330 340 350 360
Consensus ATCTACCTTT CCTAGTAGTT ATACCTGTGT GGACAACACC TCCTGGGACA CACGAATTCT CTCTGAGCAA TCTCTGTTCA ATATCTTTTA
CoxI_PCR ---------- ---------- -----TGTGT GGACAACACC TCCTGGGACA CACGAATTCT CTCTGAGCAA TCTCTGTTCA ATATCTTTTA
Illumina_1 ATCTACCTTT CCTAGTAGTT ATACCTGTGT GGACAACACC TCCTGGGACA CACGAATTCT CTCTGAGCAA TCTCTGTTCA ATATCTTTTA
Illumina_2 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_3 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_4 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_5 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_1 (5′) ATCTACCTTT CCTAGTAGTT ATACCTGTGT GGACAACACC TCCTGGGACA CACGAATTCT CTCTGAGCAA TCTCTGTTCA ATATCTTTTA
iPCR_2 (5′) ATCTACCTTT CCTAGTAGTT ATACCTGTGT GGACAACACC TCCTGGGACA CACGAATTCT CTCTGAGCAA TCTCTGTTCA ATATCTTTTA
iPCR_3 (5′) ATCTACCTTT CCTAGTAGTT ATACCTGTGT GGACAACACC TCCTGGGACA CACGAATTCT CTCTGAGCAA TCTCTGTTCA ATATCTTTTA
iPCR_4 (5′) ATCTACCTTT CCTAGTAGTT ATACCTGTGT GGACAACACC TCCTGGGACA CACGAATTCT CTCTGAGCAA TCTCTGTTCA ATATCTTTTA
iPCR_2 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_3 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_4 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

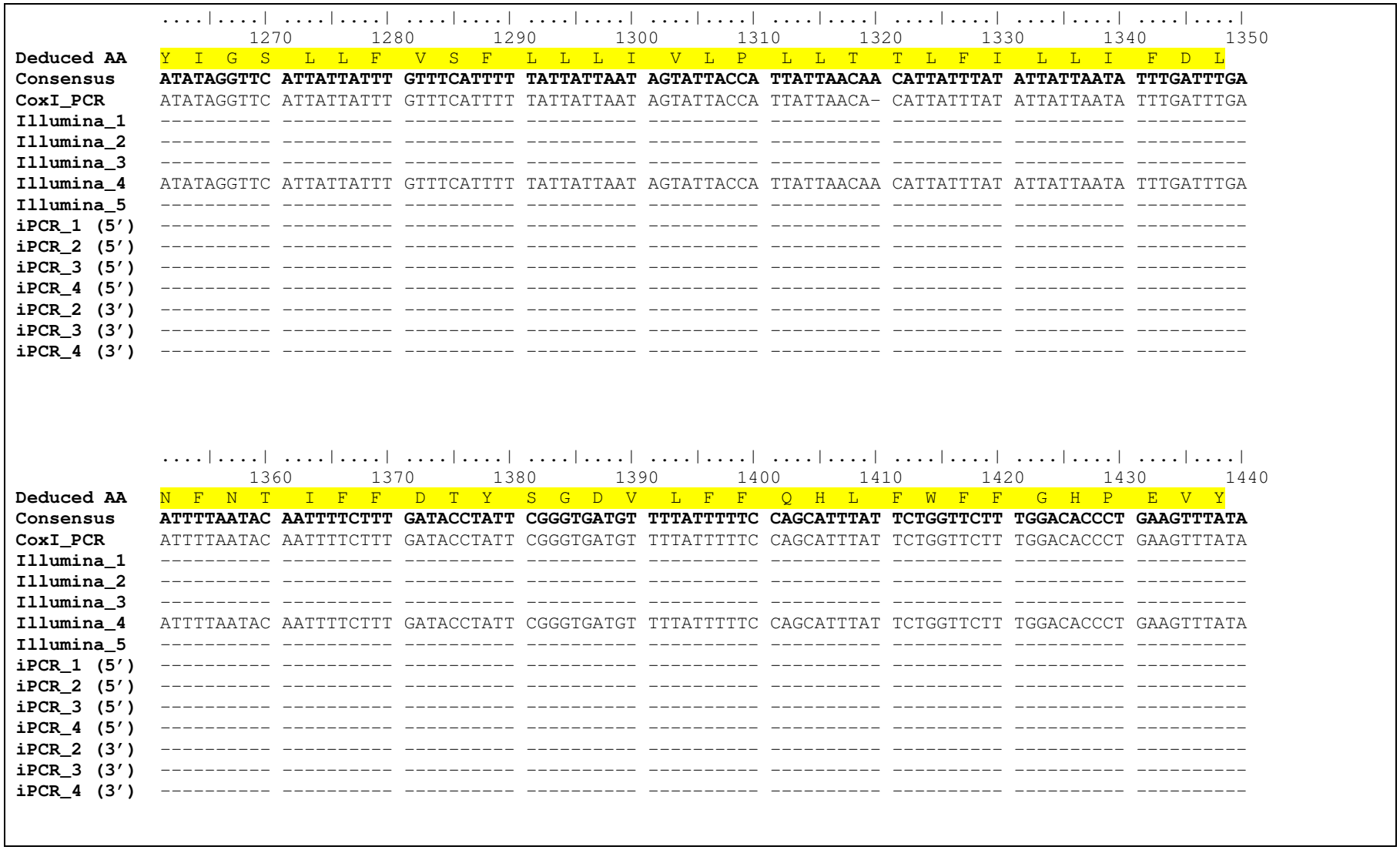

370 380 390 400 410 420 430 440 450
Consensus TTAGTTGTTG TAATATTATA ACCACAAATA CCAAATTTGG CGGCCGTGAT TATTCGGTTG AACTCCCAAA CTGCTTGTAT AAGCGGGCAA
CoxI_PCR TTAGTTGTTG TAATATTATA ACCACAAATA CCAAATTTGG CGGCCGTGAT TATTCGGTTG AACTCCCAAA CTGCTTGTAT AAGCGGGCAA
Illumina_1 TTAGTTGTTG TAATATTATA ACCACAAATA CCAAATTT-- ---------- ---------- ---------- ---------- ----------
Illumina_2 ---------- -----TTATA ACCACAAATA CCAAATTTGG CGGCCGTGAT TATTCGGTTG AACTCCCAAA CTGCTTGTAT AAGCGGGCAA
Illumina_3 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_4 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_5 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_1 (5′) TTAGTTGTTG TAATATTATA ACCACAAATA CCAAATTTGG CGGCCGTGAT TATTCGGTTG AACTCCCAAA CTGCTTGTAT AAGCGGGCAA
iPCR_2 (5′) TTAGTTGTTG TAATATTATA ACCACAAATA CCAAATTTGG CGGCCGTGAT TATTCGGTTG AACTCCCAAA CTGCTTGTAT AAGCGGGCAA
iPCR_3 (5′) TTAGTTGTTG TAATATTATA ACCACAAATA CCAAATTTGG CGGCCGTGAT TATTCGGTTG AACTCCCAAA CTGCTTGTAT AAGCGGGCAA
iPCR_4 (5′) TTAGTTGTTG TAATATTATA ACCACAAATG CCAAATTTGG CGGCCGTGAT TATTCGGTTG AACTCCCAAA CTGCTTGTAT AAGCGGGCAA
iPCR_2 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_3 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_4 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
460 470 480 490 500 510 520 530 540
Deduced AA M F T Q S E L L A M L M G K P F
Consensus TTGGTGTTGT CCAGGGAAAT AATAATCTGA CTCATCATAG TATGTTCACA CAGAGTGAGC TGCTAGCTAT GTTAATGGGA AAACCCTTTG
CoxI_PCR TTGGTGTTGT CCAGGGAAAT AATAATCTGA CTCATCATAG TATGTTCACA CAGAGTGAGC TGCTAGCTAT GTTAATGGGA AAACCCTTTG
Illumina_1 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_2 TTGGTGTTGT CCAGGGAAAT AATAATCTGA CTCATCATAG TATGTTCACA CAGAGTGAGC TGCTAGCTAT GTTAATGGGA AAACCCTTTG
Illumina_3 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_4 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_5 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_1 (5′) TTGGTGTTGT CCAGGGAGAT AATAATCTGA CTCATCATAG TATGTTCACA CAGAGTGAGC TGCTAGCTAT GTTAATGGGA AAACCCTTTG
iPCR_2 (5′) TTGGTGTTGT CCAGGGAGAT AATAATCTGA CTCATCATAG TATGTTCACA CAGAGTGAGC TGCTAGCTAT GTTAATGGGA AAACCCTTTG
iPCR_3 (5′) TTGGTGTTGT CCAGGGAGAT AATAATCTGA CTCATCATAG TATGTTCACA CAGAGTGAGC TGCTAGCTAT GTTAATGGGA AAACCCTTTG
iPCR_4 (5′) TTGGTGTTGT CCAGGGAAAT AATAATCTGA CTCATCATAG TATGTTCACA CAGAGTGAGC TGCTAGCTAT GTTAATGGGA AAACCCTTTG
iPCR_2 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_3 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_4 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

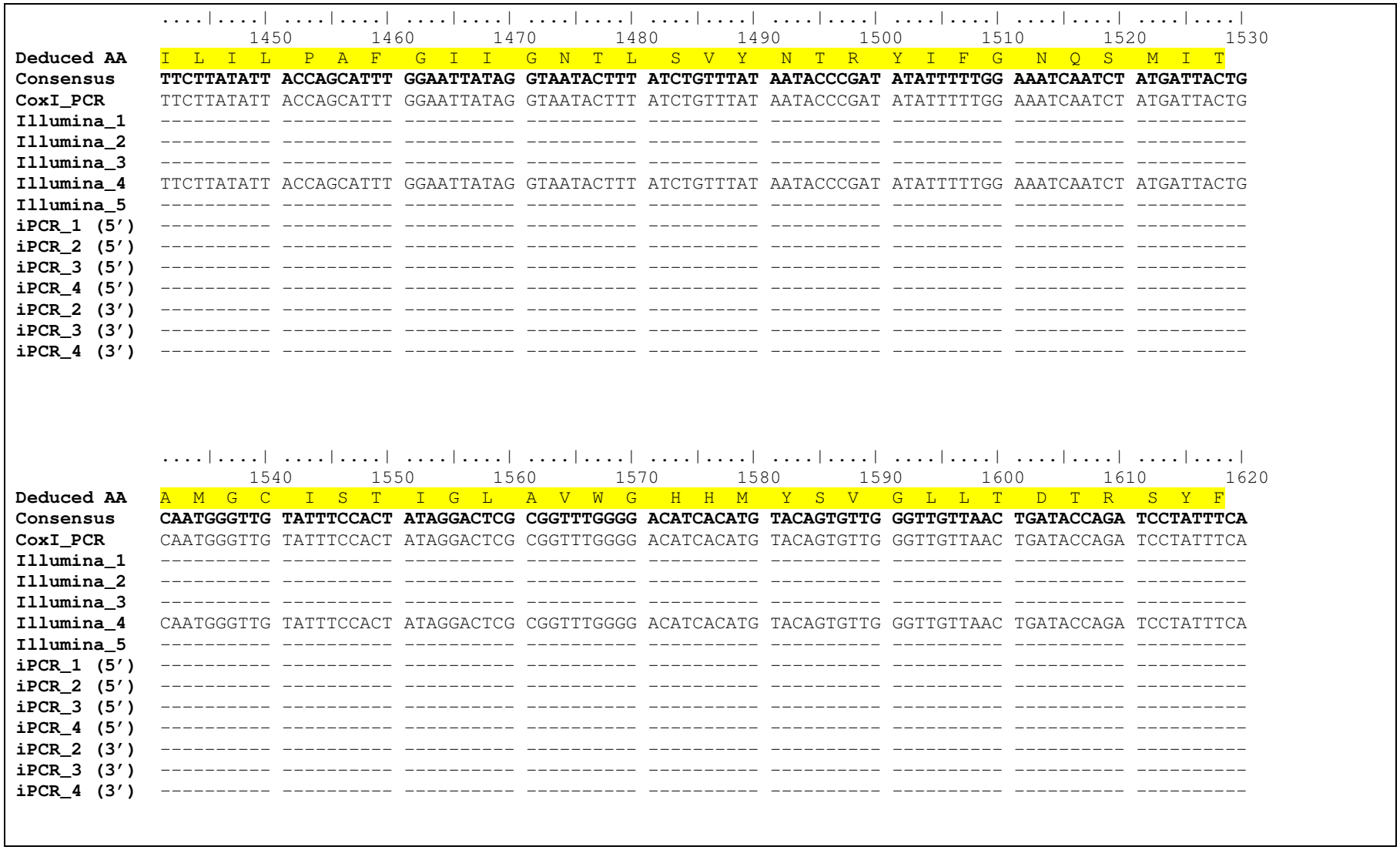
```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                        550        560        570        580        590        600        610        620        630
Deduced AA      A  D  N  T   S  M  S   W  C  K   I  N  Q  K   G  S  I   M  F  L   I  N  Y  N   E  Q  T   I  M  N
Consensus       CTGATAACAC TTCGATGAGT TGGTGTAAGA TAAATCAGAA AGGATCAATA ATGTTTCTTA TAAATTATAA TGAACAAACT ATAATGAATA
CoxI_PCR        CTGATAACAC TTCGATGAGT TGGTGTAAGA TAAATCAGAA AGGATCAATA ATGTTTCTTA TAAATTATAA TGAACAAACT ATAATGAATA
Illumina_1      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_2      CTGATAACAC TTCGATGAGT TGGTGTA--- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_3      ---------- ------GAGT TGGTGTAAGA TAAATCAGAA AGGATCAATA ATGTTTCTTA TAAATTATAA TGAACAAACT ATAATGAATA
Illumina_4      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_5      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_1 (5')     CTGATAACAC TTCGATGAGT TGGTGTAAGA TAAATCAGAA AGGATCAATA ATGTTTCTTA TAAATTATAA TGAACAAACT ATAATGAATA
iPCR_2 (5')     CTGATAACAC TTCGATGAGT TGGTGTAAGA TAAATCAGAA AGGATCAATA ATGTTTCTTA TAAATTATAA TGAACAAACT ATAATGAATA
iPCR_3 (5')     CTGATAACAC TTCGATGAGT TGGTGTAAGA TAAATCAGAG AGGATCAATA ATGTTTCTTA TAAATTATAA TGAACAAACT ATAATGAATA
iPCR_4 (5')     CTGATAACAC TTCGATGAGT TGGTGTAAGA TAAATCAGAA AGGATCAATA ATGTTTCTTA TAAATTATAA TGAACAAACT ATAATGAATA
iPCR_2 (3')     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_3 (3')     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_4 (3')     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                        640        650        660        670        680        690        700        710        720
Deduced AA      I  D  Y  N   S  N  H   F  M  D   I  F  G  F   S  Y  S   Y  N  E   F  L  Y  L   S  D  S   N  K  I
Consensus       TTGATTATAA TTCCAACCAT TTTATGGATA TATTTGGTTT TTCATATTCT TATAATGAAT TCTTATATTT ATCTGATTCA AATAAAATTA
CoxI_PCR        TTGATTATAA TTCCAACCAT TTTATGGATA TATTTGGTTT TTCATATTCT TATAATGAAT TCTTATATTT ATCTGATTCA AATAAAATTA
Illumina_1      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_2      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_3      TTGATTATAA TTCCAACCAT TTTATGGATA TATTTGGTTT TTCATATTCT TATAATGAAT TCTTATATTT ATCTGATTCA AATAAAATTA
Illumina_4      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_5      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_1 (5')     TTGATTATAA TTCCAACCAT TTTATGGATA TATTTGGTTT TTCATATTCT TATAATGAAT TCTTATATTT ATCTGATTCA AATAAAATTA
iPCR_2 (5')     TTGATTATAA TTCCAACCAT TTTATGGATA TATTTGGTTT TTCATATTCT TATAATGAAT TCTTATATTT ATCTGATTCA AATAAAATTA
iPCR_3 (5')     TTGATTATAA TTCCAACCAT TTTATGGATA TATTTGGTTT TTCATATTCT TATAATGAAT TCTTATATTT ATCTGATTCA AATAAACTTA
iPCR_4 (5')     TTGATTATAA TTCCAACCAT TTTATGGATA TATTTGGTTT TTCATATTCT TATAATGAAT TCTTATATTT ATCTGATTCA AATAAAATTA
iPCR_2 (3')     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_3 (3')     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_4 (3')     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
```

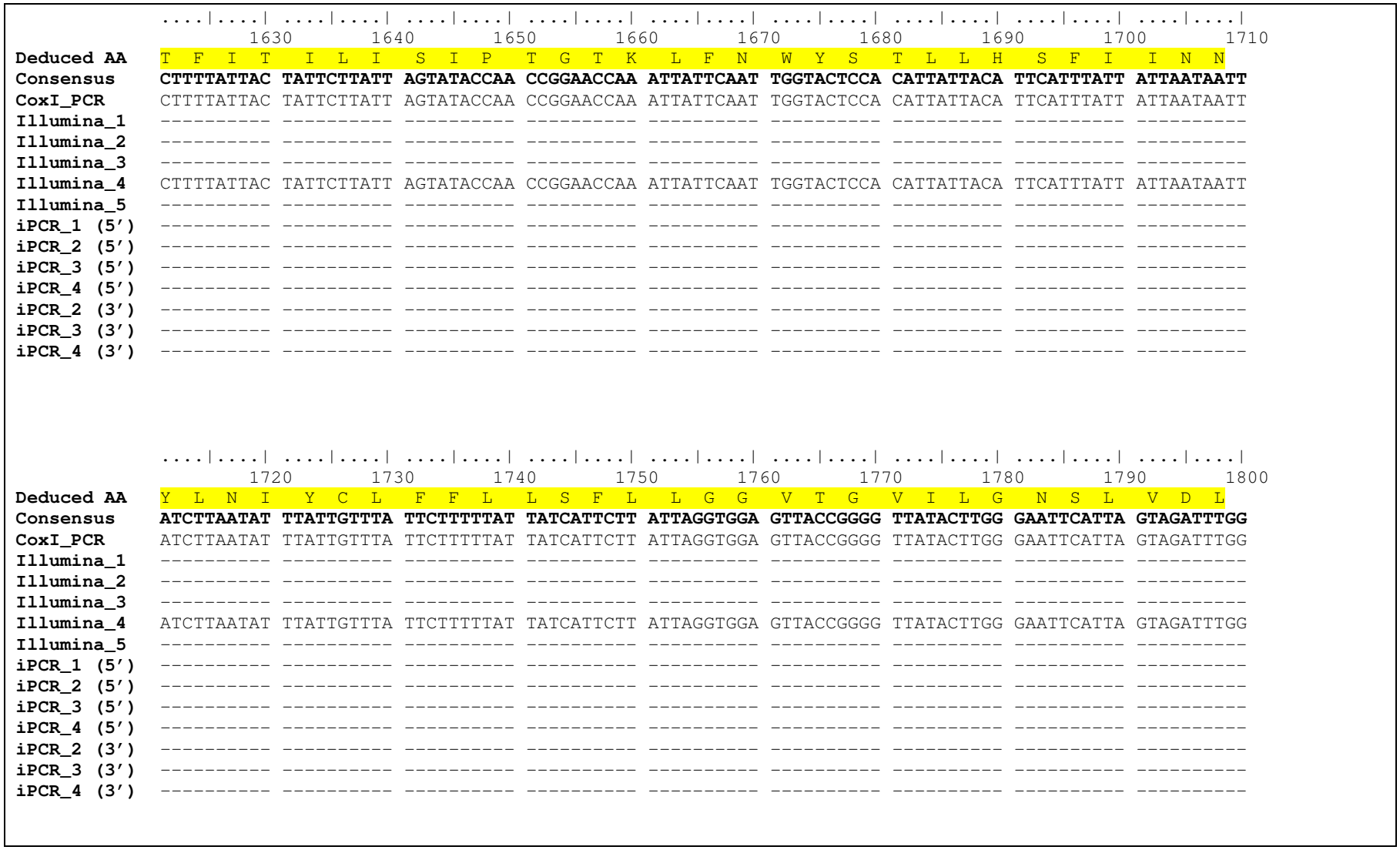

730 740 750 760 770 780 790 800 810
Deduced AA I D N I S T I N H K Y L G I Y Y L Y Y G I I F G I I G S I Y
Consensus TTGATAACAT TAGTACTATT AATCATAAAT ATTTAGGTAT ATATTATTTA TATTACGGTA TTATCTTTGG TATTATTGGA TCCATATATT
CoxI_PCR TTGATAACAT TAGTACTATT AATCATAAAT ATTTAGGTAT ATATTATTTA TATTACGGTA TTATCTTTGG TATTATTGGA TCCATATATT
Illumina_1
Illumina_2
Illumina_3 TTGATAACAT TAGTACTATT AATCATAAAT ATTTAGGTAT ATATTATTTA TATTACGGTA TTATCTTTGG TATTATTGGA TCCATATATT
Illumina_4
Illumina_5
iPCR_1 (5′) TTGATAACAT TAGTACTATT AATCATAAAT ATTTAGGTAT ATATTATTTA TATTACGGTA TTATCTTTGG TATTATTGGA TCCATATATT
iPCR_2 (5′) TTGATAACAT TAGTACTATT AATCATAGAT ATTTAGGTAT ATATTATTTA TATTACGGTA TTATCTTTGG TATGATTGGA TCCATATATT
iPCR_3 (5′) TTGATAACAT TAGTACTA--
iPCR_4 (5′) TTGATAACAT TAGTACTATT AATCATAAAT ATTTGGGTAT ATATTATTTA TATTACGGTA TTATCTTTGG TATTATTGGA TCCATATATT
iPCR_2 (3′)
iPCR_3 (3′)
iPCR_4 (3′)
820 830 840 850 860 870 880 890 900
Deduced AA S L I I R I E L G N S N I R I I N Y N Y Y N T V I T I H G L
Consensus CATTAATTAT TAGAATTGAA CTGGGTAATT CAAATATTAG GATTATCAAT TATAATTATT ATAATACTGT TATAACTATT CATGGATTAT
CoxI_PCR CATTAATTAT TAGAATTGAA CTGGGTAATT CAAATATTAG GATTATCAAT TATAATTATT ATAATACTGT TATAACTATT CATGGATTAT
Illumina_1
Illumina_2
Illumina_3 CATTAATTAT TAGAATTGAA CTGGGTAATT CAAATATTAG GATTATCAAT TATAATTATT ATAATACTGT TATAACTATT CATGGATTAT
Illumina_4 TACTGT TATAACTATT CATGGATTAT
Illumina_5
iPCR_1 (5′) CATTAATTAT TAGAATTGAA CTGGGTAATT CAAATATTAG GATTATCAAT TATAATTATT ATAATACTGT TATAACTATT CATGGATTAT
iPCR_2 (5′) CATTAATTAT TAGAATTGAA CTGGGTAATT CAAATATTAG GATTATCAAT TATAATTATT ATAATACTGT TATAACTATT CATGGATAAT
iPCR_3 (5′)
iPCR_4 (5′) CATTAATTAT TAGAATTGAA CTGGGTAATT CAAATATTAG GATTATCAAT TATAATTATT ATAATACTGT TATAACTATT CATGGATTAT
iPCR_2 (3′)
iPCR_3 (3′)
iPCR_4 (3′)

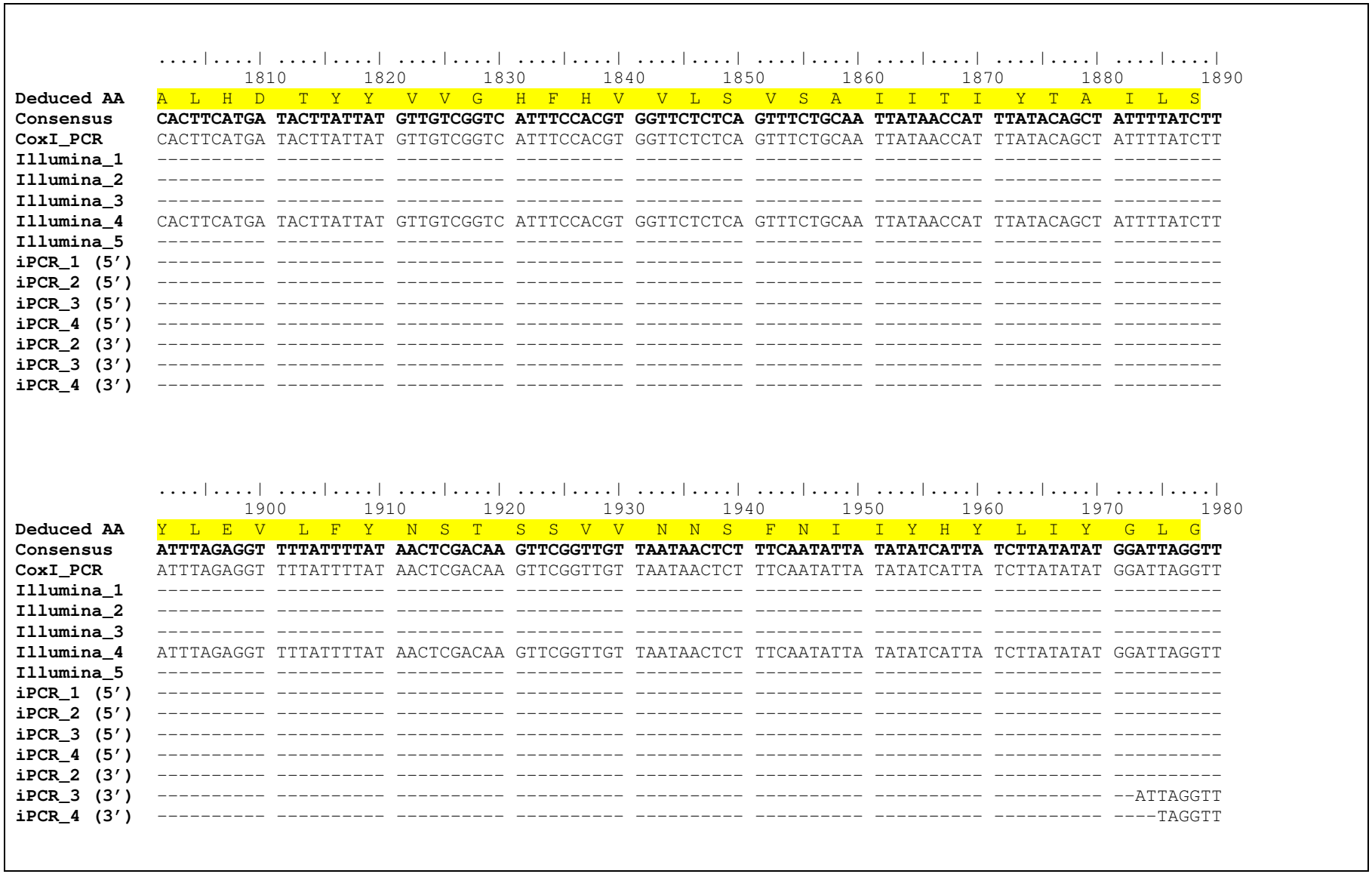

910 920 930 940 950 960 970 980 990
Deduced AA L M I F F L I I P V L F S G L G N Y I S P I N I I K M E I S
Consensus TAATGATCTT TTTTCTTATT ATTCCAGTAT TATTTAGTGG ATTAGGAAAT TATATTTCTC CTATTAATAT TATCAAAATG GAAATATCCC
CoxI_PCR TAATGATCTT TTTTCTTATT ATTCCAGTAT TATTTAGTGG ATTAGGAAAT TATATTTCTC CTATTAATAT TATCAAAATG GAAATATCCC
Illumina_1
Illumina_2
Illumina_3 TAATGATCTT TTTTCTTATT ATTCCAGTAT TATT------
Illumina_4 TAATGATCTT TTTTCTTATT ATTCCAGTAT TATTTAGTGG ATTAGGAAAT TATATTTCTC CTATTAATAT TATCAAAATG GAAATATCCC
Illumina_5
iPCR_1 (5′) TAATGATCTT TTTTCTTATT ATTCCAGTAT TATTTA----
iPCR_2 (5′) TAACGATCCT TTGTCCTAAT AATCCAGTAT TATTTT----
iPCR_3 (5′)
iPCR_4 (5′) TAAGGATCTT TTTTCTTATT ATTCCAGTAT TATTTA----
iPCR_2 (3′)
iPCR_3 (3′)
iPCR_4 (3′)
1000 1010 1020 1030 1040 1050 1060 1070 1080
Deduced AA L P R V N N L S L I L L L F S F I I F I F N N L I D F G V G
Consensus TTCCGAGAGT GAATAATCTT TCATTAATAT TATTATTATT TTCATTTATT ATTTTTATCT TTAATAATCT TATAGATTTT GGTGTTGGTT
CoxI_PCR TTCCGAGAGT GAATAATCTT TCATTAATAT TATTATTATT TTCATTTAT- ATTTTTATCT TTAATAATCT TATAGATTTT GGTGTTGGTT
Illumina_1
Illumina_2
Illumina_3
Illumina_4 TTCCGAGAGT GAATAATCTT TCATTAATAT TATTATTATT TTCATTTATT ATTTTTATCT TTAATAATCT TATAGATTTT GGTGTTGGTT
Illumina_5
iPCR_1 (5′)
iPCR_2 (5′)
iPCR_3 (5′)
iPCR_4 (5′)
iPCR_2 (3′)
iPCR_3 (3′)
iPCR_4 (3′)

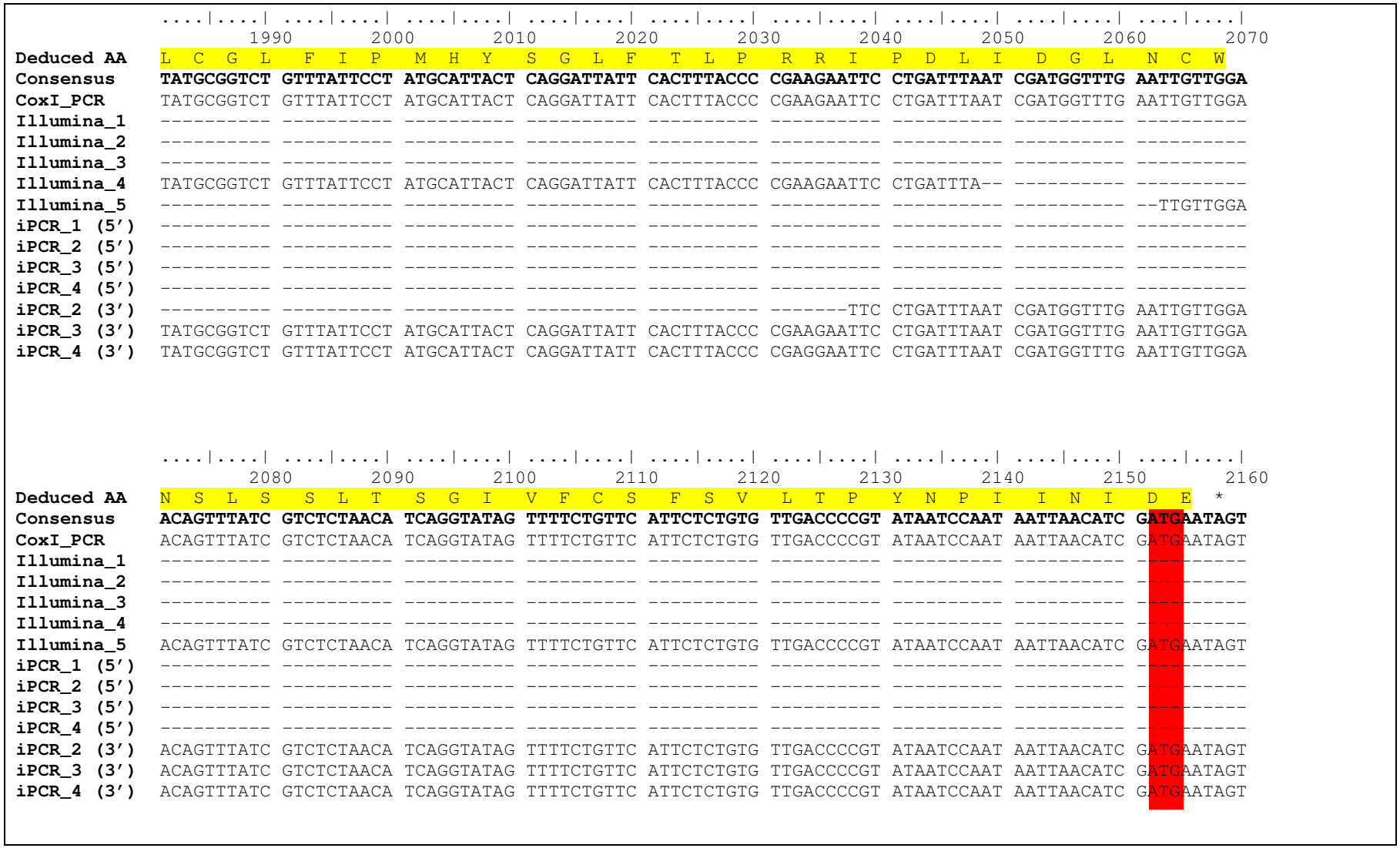

....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
1090 1100 1110 1120 1130 1140 1150 1160 1170
Deduced AA L G W T L Y P P L S T I I Y L L S I N I I I I G L L V S G I
Consensus TAGGTTGGAC CCTATACCCT CCATTATCCA CTATTATATA TCTTTTATCG ATCAATATAA TAATTATTGG TTTATTGGTT TCTGGAATTT
CoxI_PCR TAGGTTGGAC CCTATACCCT CCATTATCCA CTATTATATA TCTTTTATCG ATCAATATAA TAATTATTGG TTTATTGGTT TCTGGAATTT
Illumina_1 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_2 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_3 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_4 TAGGTTGGAC CCTATACCCT CCATTATCCA CTATTATATA TCTTTTATCG ATCAATATAA TAATTATTGG TTTATTGGTT TCTGGAATTT
Illumina_5 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_1 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_2 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_3 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_4 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_2 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_3 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_4 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
1180 1190 1200 1210 1220 1230 1240 1250 1260
Deduced AA S S I L S S I N F I Y T I I I I Y T G L N N S D I N M Y S T
Consensus CCAGTATATT ATCTAGTATT AATTTTATTT ATACTATTAT TATTATTTAT ACAGGATTAA ATAATTCCGA TATTAATATG TATTCAACGT
CoxI_PCR CCAGTATATT ATCTAGTATT AATTTTATTT ATACTATTAT TATTATTTAT ACAGGATTAA ATAATTCCGA TATTAATATG TATTCAACGT
Illumina_1 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_2 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_3 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_4 CCAGTATATT ATCTAGTATT AATTTTATTT ATACTATTAT TATTATTTAT ACAGGATTAA ATAATTCCGA TATTAATATG TATTCAACGT
Illumina_5 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_1 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_2 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_3 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_4 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_2 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_3 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_4 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------

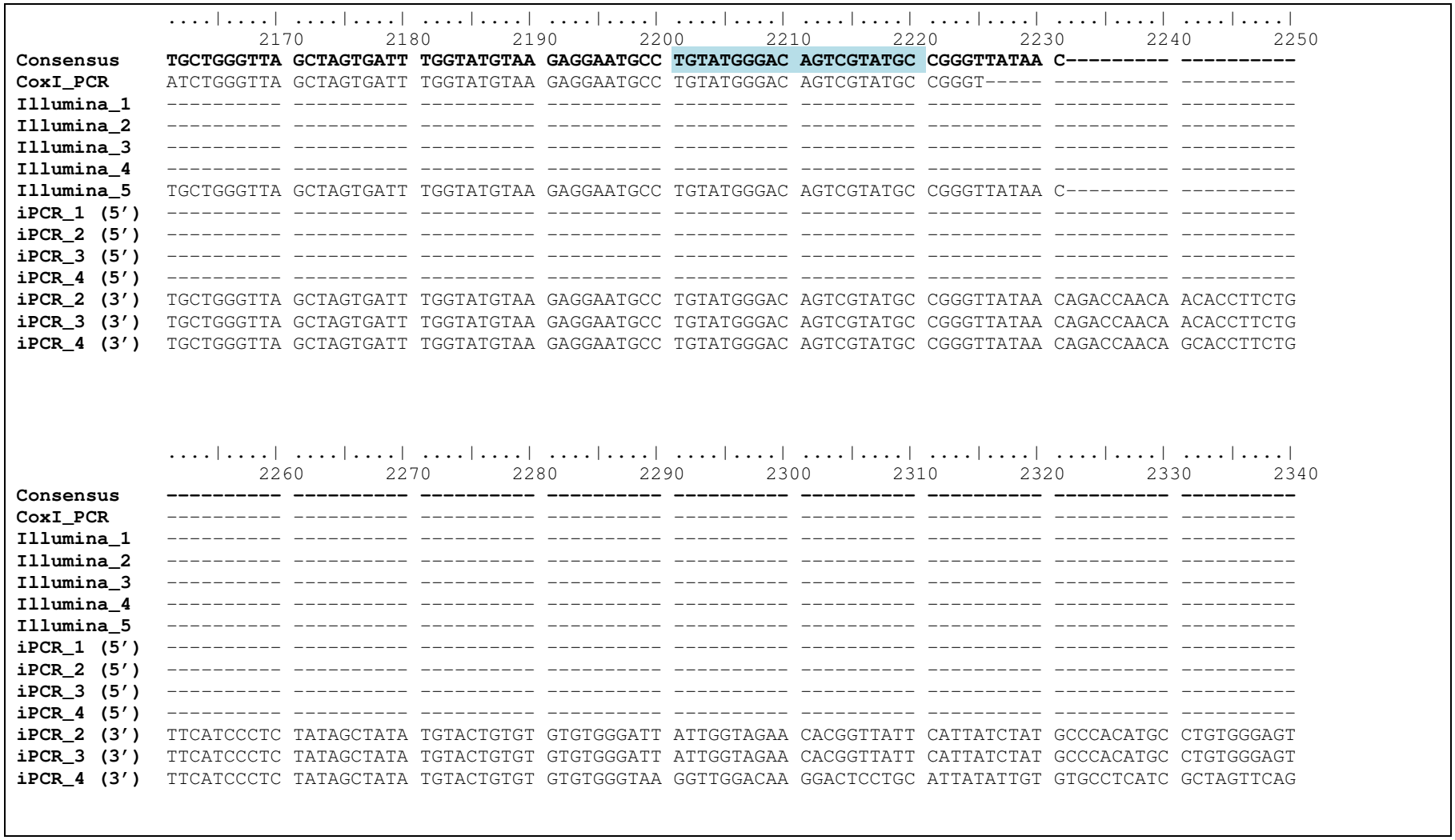

....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
1270 1280 1290 1300 1310 1320 1330 1340 1350
Deduced AA Y I G S L L F V S F L L L I V L P L L T T L F I L L I F D L
Consensus ATATAGGTTC ATTATTATTT GTTTCATTTT TATTATTAAT AGTATTACCA TTATTAACAA CATTATTTAT ATTATTAATA TTTGATTTGA
CoxI_PCR ATATAGGTTC ATTATTATTT GTTTCATTTT TATTATTAAT AGTATTACCA TTATTAACA- CATTATTTAT ATTATTAATA TTTGATTTGA
Illumina_1 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_2 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_3 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_4 ATATAGGTTC ATTATTATTT GTTTCATTTT TATTATTAAT AGTATTACCA TTATTAACAA CATTATTTAT ATTATTAATA TTTGATTTGA
Illumina_5 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_1 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_2 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_3 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_4 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_2 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_3 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_4 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
1360 1370 1380 1390 1400 1410 1420 1430 1440
Deduced AA N F N T I F F D T Y S G D V L F F Q H L F W F F G H P E V Y
Consensus ATTTTAATAC AATTTTCTTT GATACCTATT CGGGTGATGT TTTATTTTTC CAGCATTTAT TCTGGTTCTT TGGACACCCT GAAGTTTATA
CoxI_PCR ATTTTAATAC AATTTTCTTT GATACCTATT CGGGTGATGT TTTATTTTTC CAGCATTTAT TCTGGTTCTT TGGACACCCT GAAGTTTATA
Illumina_1 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_2 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_3 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_4 ATTTTAATAC AATTTTCTTT GATACCTATT CGGGTGATGT TTTATTTTTC CAGCATTTAT TCTGGTTCTT TGGACACCCT GAAGTTTATA
Illumina_5 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_1 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_2 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_3 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_4 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_2 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_3 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_4 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------


1450 1460 1470 1480 1490 1500 1510 1520 1530
Deduced AA I L I L P A F G I I G N T L S V Y N T R Y I F G N Q S M I T
Consensus TTCTTATATT ACCAGCATTT GGAATTATAG GTAATACTTT ATCTGTTTAT AATACCCGAT ATATTTTTGG AAATCAATCT ATGATTACTG
CoxI_PCR TTCTTATATT ACCAGCATTT GGAATTATAG GTAATACTTT ATCTGTTTAT AATACCCGAT ATATTTTTGG AAATCAATCT ATGATTACTG
Illumina_1
Illumina_2
Illumina_3
Illumina_4 TTCTTATATT ACCAGCATTT GGAATTATAG GTAATACTTT ATCTGTTTAT AATACCCGAT ATATTTTTGG AAATCAATCT ATGATTACTG
Illumina_5
iPCR_1 (5′)
iPCR_2 (5′)
iPCR_3 (5′)
iPCR_4 (5′)
iPCR_2 (3′)
iPCR_3 (3′)
iPCR_4 (3′)
1540 1550 1560 1570 1580 1590 1600 1610 1620
Deduced AA A M G C I S T I G L A V W G H H M Y S V G L L T D T R S Y F
Consensus CAATGGGTTG TATTTCCACT ATAGGACTCG CGGTTTGGGG ACATCACATG TACAGTGTTG GGTTGTTAAC TGATACCAGA TCCTATTTCA
CoxI_PCR CAATGGGTTG TATTTCCACT ATAGGACTCG CGGTTTGGGG ACATCACATG TACAGTGTTG GGTTGTTAAC TGATACCAGA TCCTATTTCA
Illumina_1
Illumina_2
Illumina_3
Illumina_4 CAATGGGTTG TATTTCCACT ATAGGACTCG CGGTTTGGGG ACATCACATG TACAGTGTTG GGTTGTTAAC TGATACCAGA TCCTATTTCA
Illumina_5
iPCR_1 (5′)
iPCR_2 (5′)
iPCR_3 (5′)
iPCR_4 (5′)
iPCR_2 (3′)
iPCR_3 (3′)
iPCR_4 (3′)


....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
1630 1640 1650 1660 1670 1680 1690 1700 1710
Deduced AA T F I T I L I S I P T G T K L F N W Y S T L L H S F I I N N
Consensus CTTTTATTAC TATTCTTATT AGTATACCAA CCGGAACCAA ATTATTCAAT TGGTACTCCA CATTATTACA TTCATTTATT ATTAATAATT
CoxI_PCR CTTTTATTAC TATTCTTATT AGTATACCAA CCGGAACCAA ATTATTCAAT TGGTACTCCA CATTATTACA TTCATTTATT ATTAATAATT
Illumina_1 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_2 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_3 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_4 CTTTTATTAC TATTCTTATT AGTATACCAA CCGGAACCAA ATTATTCAAT TGGTACTCCA CATTATTACA TTCATTTATT ATTAATAATT
Illumina_5 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_1 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_2 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_3 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_4 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_2 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_3 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_4 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
1720 1730 1740 1750 1760 1770 1780 1790 1800
Deduced AA Y L N I Y C L F F L L S F L L G G V T G V I L G N S L V D L
Consensus ATCTTAATAT TTATTGTTTA TTCTTTTTAT TATCATTCTT ATTAGGTGGA GTTACCGGGG TTATACTTGG GAATTCATTA GTAGATTTGG
CoxI_PCR ATCTTAATAT TTATTGTTTA TTCTTTTTAT TATCATTCTT ATTAGGTGGA GTTACCGGGG TTATACTTGG GAATTCATTA GTAGATTTGG
Illumina_1 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_2 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_3 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_4 ATCTTAATAT TTATTGTTTA TTCTTTTTAT TATCATTCTT ATTAGGTGGA GTTACCGGGG TTATACTTGG GAATTCATTA GTAGATTTGG
Illumina_5 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_1 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_2 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_3 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_4 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_2 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_3 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_4 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------


1810 1820 1830 1840 1850 1860 1870 1880 1890
Deduced AA A L H D T Y Y V V G H F H V V L S V S A I I T I Y T A I L S
Consensus CACTTCATGA TACTTATTAT GTTGTCGGTC ATTTCCACGT GGTTCTCTCA GTTTCTGCAA TTATAACCAT TTATACAGCT ATTTTATCTT
CoxI_PCR CACTTCATGA TACTTATTAT GTTGTCGGTC ATTTCCACGT GGTTCTCTCA GTTTCTGCAA TTATAACCAT TTATACAGCT ATTTTATCTT
Illumina_1 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_2 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_3 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_4 CACTTCATGA TACTTATTAT GTTGTCGGTC ATTTCCACGT GGTTCTCTCA GTTTCTGCAA TTATAACCAT TTATACAGCT ATTTTATCTT
Illumina_5 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_1 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_2 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_3 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_4 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_2 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_3 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_4 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
1900 1910 1920 1930 1940 1950 1960 1970 1980
Deduced AA Y L E V L F Y N S T S S V V N N S F N I I Y H Y L I Y G L G
Consensus ATTTAGAGGT TTTATTTTAT AACTCGACAA GTTCGGTTGT TAATAACTCT TTCAATATTA TATATCATTA TCTTATATAT GGATTAGGTT
CoxI_PCR ATTTAGAGGT TTTATTTTAT AACTCGACAA GTTCGGTTGT TAATAACTCT TTCAATATTA TATATCATTA TCTTATATAT GGATTAGGTT
Illumina_1 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_2 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_3 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_4 ATTTAGAGGT TTTATTTTAT AACTCGACAA GTTCGGTTGT TAATAACTCT TTCAATATTA TATATCATTA TCTTATATAT GGATTAGGTT
Illumina_5 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_1 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_2 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_3 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_4 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_2 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_3 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- --ATTAGGTT
iPCR_4 (3′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----TAGGTT


....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
1990 2000 2010 2020 2030 2040 2050 2060 2070
Deduced AA L C G L F I P M H Y S G L F T L P R R I P D L I D G L N C W
Consensus TATGCGGTCT GTTTATTCCT ATGCATTACT CAGGATTATT CACTTTACCC CGAAGAATTC CTGATTTAAT CGATGGTTTG AATTGTTGGA
CoxI_PCR TATGCGGTCT GTTTATTCCT ATGCATTACT CAGGATTATT CACTTTACCC CGAAGAATTC CTGATTTAAT CGATGGTTTG AATTGTTGGA
Illumina_1 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_2 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_3 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_4 TATGCGGTCT GTTTATTCCT ATGCATTACT CAGGATTATT CACTTTACCC CGAAGAATTC CTGATTTA-- ---------- ----------
Illumina_5 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- --TTGTTGGA
iPCR_1 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_2 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_3 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_4 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_2 (3′) ---------- ---------- ---------- ---------- ---------- -------TTC CTGATTTAAT CGATGGTTTG AATTGTTGGA
iPCR_3 (3′) TATGCGGTCT GTTTATTCCT ATGCATTACT CAGGATTATT CACTTTACCC CGAAGAATTC CTGATTTAAT CGATGGTTTG AATTGTTGGA
iPCR_4 (3′) TATGCGGTCT GTTTATTCCT ATGCATTACT CAGGATTATT CACTTTACCC CGAGGAATTC CTGATTTAAT CGATGGTTTG AATTGTTGGA
....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
2080 2090 2100 2110 2120 2130 2140 2150 2160
Deduced AA N S L S S L T S G I V F C S F S V L T P Y N P I I N I D E *
Consensus ACAGTTTATC GTCTCTAACA TCAGGTATAG TTTTCTGTTC ATTCTCTGTG TTGACCCCGT ATAATCCAAT AATTAACATC GATGAATAGT
CoxI_PCR ACAGTTTATC GTCTCTAACA TCAGGTATAG TTTTCTGTTC ATTCTCTGTG TTGACCCCGT ATAATCCAAT AATTAACATC GATGAATAGT
Illumina_1 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_2 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_3 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_4 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_5 ACAGTTTATC GTCTCTAACA TCAGGTATAG TTTTCTGTTC ATTCTCTGTG TTGACCCCGT ATAATCCAAT AATTAACATC GATGAATAGT
iPCR_1 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_2 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_3 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_4 (5′) ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_2 (3′) ACAGTTTATC GTCTCTAACA TCAGGTATAG TTTTCTGTTC ATTCTCTGTG TTGACCCCGT ATAATCCAAT AATTAACATC GATGAATAGT
iPCR_3 (3′) ACAGTTTATC GTCTCTAACA TCAGGTATAG TTTTCTGTTC ATTCTCTGTG TTGACCCCGT ATAATCCAAT AATTAACATC GATGAATAGT
iPCR_4 (3′) ACAGTTTATC GTCTCTAACA TCAGGTATAG TTTTCTGTTC ATTCTCTGTG TTGACCCCGT ATAATCCAAT AATTAACATC GATGAATAGT


2170 2180 2190 2200 2210 2220 2230 2240 2250
Consensus TGCTGGGTTA GCTAGTGATT TGGTATGTAA GAGGAATGCC TGTATGGGAC AGTCGTATGC CGGGTTATAA C
CoxI_PCR ATCTGGGTTA GCTAGTGATT TGGTATGTAA GAGGAATGCC TGTATGGGAC AGTCGTATGC CGGGT
Illumina_1
Illumina_2
Illumina_3
Illumina_4
Illumina_5 TGCTGGGTTA GCTAGTGATT TGGTATGTAA GAGGAATGCC TGTATGGGAC AGTCGTATGC CGGGTTATAA C
iPCR_1 (5′)
iPCR_2 (5′)
iPCR_3 (5′)
iPCR_4 (5′)
iPCR_2 (3′) TGCTGGGTTA GCTAGTGATT TGGTATGTAA GAGGAATGCC TGTATGGGAC AGTCGTATGC CGGGTTATAA CAGACCAACA ACACCTTCTG
iPCR_3 (3′) TGCTGGGTTA GCTAGTGATT TGGTATGTAA GAGGAATGCC TGTATGGGAC AGTCGTATGC CGGGTTATAA CAGACCAACA ACACCTTCTG
iPCR_4 (3′) TGCTGGGTTA GCTAGTGATT TGGTATGTAA GAGGAATGCC TGTATGGGAC AGTCGTATGC CGGGTTATAA CAGACCAACA GCACCTTCTG
2260 2270 2280 2290 2300 2310 2320 2330 2340
Consensus
CoxI_PCR
Illumina_1
Illumina_2
Illumina_3
Illumina_4
Illumina_5
iPCR_1 (5′)
iPCR_2 (5′)
iPCR_3 (5′)
iPCR_4 (5′)
iPCR_2 (3′) TTCATCCCTC TATAGCTATA TGTACTGTGT GTGTGGGATT ATTGGTAGAA CACGGTTATT CATTATCTAT GCCCACATGC CTGTGGGAGT
iPCR_3 (3′) TTCATCCCTC TATAGCTATA TGTACTGTGT GTGTGGGATT ATTGGTAGAA CACGGTTATT CATTATCTAT GCCCACATGC CTGTGGGAGT
iPCR_4 (3′) TTCATCCCTC TATAGCTATA TGTACTGTGT GTGTGGGTAA GGTTGGACAA GGACTCCTGC ATTATATTGT GTGCCTCATC GCTAGTTCAG

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                        2350       2360       2370       2380       2390       2400       2410       2420       2430
Consensus       ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
CoxI_PCR        ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_1      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_2      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_3      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_4      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_5      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_1 (5′)     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_2 (5′)     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_3 (5′)     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_4 (5′)     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_2 (3′)     GTTGGATAAA TATTAGAGTT GAACACACAC AGATAACACA CACAGATAAT ATTAATAGAA CGGTTATACG TGGGTTGGTA GTTCAGATAT
iPCR_3 (3′)     GTTGGATAAA TATTAGAGTT GAACACACAC AGATAACACA CACAGATAAT ATTAATAGAA CGGTTATACG TGGGTTGGTA GTTCAGATAT
iPCR_4 (3′)     ATATCTACCT TTCCTAGTAG TTATACCTGT GTGGACAACA CCTCCTGGGA CACACGAATT CTCTCTGAGC AATCTCTGTT CAATATCTTT
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                        2440       2450       2460       2470       2480       2490       2500       2510       2520
Consensus       ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
CoxI_PCR        ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_1      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_2      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_3      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_4      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_5      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_1 (5′)     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_2 (5′)     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_3 (5′)     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_4 (5′)     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_2 (3′)     CTACCTTTCC TAGTAGTTAT ACCTGTGTGG ACAACACCTC CTGGGACACA CGAATTCTCT CTGAGCAATC TCTGTTCAAT ATCTTTTATT
iPCR_3 (3′)     CTACCTTTCC TAGTAGTTAT ACCTGTGTGG ACAACACCTC CTGGGACACA CGAATTCTCT CTGAGCAATC TCTGTTCAAT ATCTTTTATT
iPCR_4 (3′)     TATTAGTTGT TGTAATATTA TAACCACAAA TGCCAAATTT GGCGGCCGTG ATTATTCGGT TGAACTCCCA AACTGCTTGT ATAAGCGGGC
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                        2530       2540       2550       2560       2570       2580       2590       2600       2610
Consensus       ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
CoxI_PCR        ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_1      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_2      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_3      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_4      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Illumina_5      ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_1 (5')     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_2 (5')     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_3 (5')     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_4 (5')     ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
iPCR_2 (3')     AGTTGTTGTA ATATTATAAC CACAAATACC AAATTTGGCG GCCGTGATTA TTCGGTTGAA CTCCCAAACT GCTTGTATAA GCGGGCAATT
iPCR_3 (3')     AGTTGTTGTA ATATTATAAC CACAAATACC AAATTTGGCG GCCGTGATTA TTCGGTTGAA CTCCCAAACT GCTTGTATAA GCGGGCAATT
iPCR_4 (3')     AATTGGTGTT GTCCAGGGAA ATAATAATCT GACTCATCAT AGT------- ---------- ---------- ---------- ----------

                ....|....| ....|....| ....|....| ....|....
                        2620       2630       2640
Consensus       ---------- ---------- ---------- ---------
CoxI_PCR        ---------- ---------- ---------- ---------
Illumina_1      ---------- ---------- ---------- ---------
Illumina_2      ---------- ---------- ---------- ---------
Illumina_3      ---------- ---------- ---------- ---------
Illumina_4      ---------- ---------- ---------- ---------
Illumina_5      ---------- ---------- ---------- ---------
iPCR_1 (5')     ---------- ---------- ---------- ---------
iPCR_2 (5')     ---------- ---------- ---------- ---------
iPCR_3 (5')     ---------- ---------- ---------- ---------
iPCR_4 (5')     ---------- ---------- ---------- ---------
iPCR_2 (3')     GGTGTTGTCC AGGGAGATAA TAATCTGACT CATCATAGT
iPCR_3 (3')     GGTGTTGTCC AGGGAGATAA TAATCTGACT CATCATAGT
iPCR_4 (3')     ---------- ---------- ---------- ---------
```

Figure S14: Alignment of the mitochondrial genome fragments from *Chromera velia* CCAP 1602/1 based on five genomic nodes of the metagenome sequencing project according to Petersen et al. (2014) and four sequences obtained by inverse PCR amplifications (iPCR_1 to iPCR_4). Color code: yellow = deduced aa sequence of *cox*I; green = start codon; red = stop codon; light gray = overlapping Illumina sequences; light green = homologous regions Illumina/iPCR; light red = homologous regions iPCRs; light blue = primer sequences to amplify the complete *coxI* gene. The nucleotide consensus sequence of all genome and iPCR sequences which is shown in bold is deposited under the accession number KC899110.

Table S7: Mitochondrial metabolic pathway inventory according to Danne et al. (2013) adding transcriptomic and genomic information of my new established datasets for *Perkinsus olseni* (Po), *Prorocentrum minimum* (Prom), *Chromera velia* (Cv) and *Vitrella brassicaformis* (Vb). The new added columns are marked by yellow. Abbreviations: Tet = Tetrahymena, Pm = *Perkinsus* marinus, Dino = Dinoflagellata, Cp = *Cryptosporidium parvum*, Tg = *Toxoplasma gondii*, Plas/Pir = *Plasmodium*/Piroplasmida; Indices: a/b = subunit a/b; f = fusion protein; ID = identity uncertain; n = nuclear-encoded; P = located in plastid; ? = location uncertain; * = absent from piroplasms

Function	Protein	Abbreviation	Organism									
			Ciliates	Dinozoa				Apicomplexa				
			Tet	*Pm*	*Po*	Dino	*Prom*	*Cv*	*Vb*	*Cp*	*Tg*	Plas/Pir
PDH complex	Pyruvat Dehydrogenase E1 alpha	PDHE1-A	✓	-	-	-	-	-	-	-	-	-
	Pyruvat Dehydrogenase E1 beta	PDHE1-B	✓	-	-	-	-	✓^{n}	✓^{n}	-	-	-
	Pyruvat Dehydrogenase E2	PDHC	✓	-	-	-	-	✓^{n}	✓^{n}	-	-	-
	KADH E3 (all KADHs)	KADH3	✓	✓	✓^{n}	✓	✓	✓^{n}	✓^{n}	-	✓	✓
BCKDH complex	Branched chain ketoacid dehydrogenase E1 alpha	BCKDE1A	✓	✓	✓^{n}	✓	✓	✓^{n}	✓^{n}	-	✓	✓
	Branched chain ketoacid dehydrogenase E1 beta	BCKDE1B	✓	✓	✓^{n}	✓	✓	✓^{n}	✓^{n}	-	✓	✓
	Branched chain ketoacid dehydrogenase E2	BCKADE2	✓	✓	✓^{n}	✓	✓	✓^{n}	✓^{n}	-	✓	✓
	KADH E3 (all KADHs)	KADH3	✓	✓	✓^{n}	✓	✓	✓^{n}	✓^{n}	-	✓	✓
Lipoic acid	Lipoate protein ligase	LipL	✓	✓	✓^{n}	✓	✓	✓^{n}	✓^{n}	-	✓	✓
Degradation of BCAA	Branched chain amino acid transaminase	BCAT	✓	✓	✓^{n}	✓	✓	✓^{n}	✓^{n}	-	✓	-
	Acyl-coA dehydrogenase	ACAD	✓	✓	✓^{n}	✓	✓	✓^{n}	✓^{n}	-	✓	-
	3-Hydroxyisobutyryl-coA hydrolase	HIBCH	✓	✓	✓^{n}	✓	✓	✓^{n}	✓^{n}	-	✓	-
	3-Hydroxyisobutyrate-coA dehydrogenase	HIBADH	✓	✓	✓^{n}	✓	✓	✓^{n}	✓^{n}	-	✓	-
	3-Hydroxyacyl-coA dehydrogenase	HADH	-	✓	✓^{n}	✓	✓	-	✓^{n}	-	✓	-
	Methylmalonate-semi-aldehyde dehydrogenase	MMSDH	✓	✓	✓^{n}	✓	✓	✓^{n}	✓^{n}	-	✓	-
	Isovaleryl-CoA dehydrogenase	IVD	✓	✓	✓^{n}	✓	✓	✓^{n}	✓^{n}	-	✓	-
	Hydroxymethylglutaryl-CoA lyase	HMGCL	✓	-	-	✓	-	✓^{n}	✓^{n}	-	✓	-
	Acetyl-CoA C-acyltransferase	ACAT	✓	✓	✓^{n}	✓	✓	✓^{n}	✓^{n}	-	✓	-
2-Methylcitrate cycle	2-Methyl citrate synthase	MCS	✓	✓	✓^{n}	✓	✓	✓^{n}	✓^{n}	-	✓	-
	Propionyl-CoA synthetase	PrpE	✓	-	-	✓	✓	✓^{n}	✓^{n}	-	✓	-
	2-Methyl citrate dehydratase	MCD	✓	✓	✓^{n}	✓	✓	✓^{n}	✓^{n}	-	✓	-
	2-Methylisocitrate lyase	MICL	✓	✓	✓^{n}	✓	✓	✓^{n}	✓^{n}	-	✓	-

Function	Protein	Abbr.	Organism									
			Ciliates	Dinozoa				Apicomplexa				
			Tet	*Pm*	*Po*	Dino	*Prom*	*Cv*	*Vb*	*Cp*	*Tg*	Plas/Pir
Fatty acid β-oxidation	Acyl-CoA dehydrogenase	ACAD	✓	✓	✓[n]	✓	✓	✓[n]	✓[n]	-	✓	-
	Enoyl-CoA hydratase	ECH	✓	✓	✓[n]	✓	✓	✓[n]	✓[n]	-	✓	-
	Hydroxyacyl-CoA dehydrogenase	HADH	✓	✓	✓[n]	✓	✓	-	✓[n]	-	✓	-
	Ketoacyl-CoA thiolase	KAT	✓	✓	✓[n]	✓	✓	✓[n]	✓[n]	-	✓	-
TCA cycle	Citrate synthase	CS	✓	✓	✓[n]	✓	✓	✓[n]	✓[n]	-	✓	✓
	Aconitase	ACO2	✓	✓	✓[n]	✓	✓	✓[n]	✓[n]	-	✓	✓
	Isocitrate dehydrogenase NAD(H)-dependent	IDH1	✓	-	-	-	-	-	-	-	-	-
	Isocitrate dehydrogenase NADP(H)-dependent	IDH2	✓	✓	✓[n]	✓	✓	✓[n]	✓[n]	-	✓	✓
	NAD(P)+ transhydrogenase	pntA	✓	✓	✓[n]	✓	✓	✓[n]	✓[n]	✓[*]	✓	✓
	Ketogluatarate dehydrogenase E1	OGDH	✓	-	-	✓	✓	✓[n]	✓[n]	-	✓	✓
	Ketoglutarate dehydrogenase E2	KGD2	✓	-	-	✓	✓	✓[n]	✓[n]	-	✓	✓
	KADH E3 (shared by all KADHs)	KADH3	✓	✓	✓[n]	✓	✓	✓[n]	✓[n]	-	✓	✓
	Succinyl-CoA synthetase alpha	sucD	✓	✓	✓[n]	✓	✓	✓[n]	✓[n]	-	✓	✓
	Succinyl-CoA synthetase beta	sucC	✓	✓	✓[n]	✓	✓	✓[n]	✓[n]	-	✓	✓
	Succinate dehydrogenase Fe-S	sdhB	✓	✓	✓[n]	✓	✓	✓[n]	✓[n]	-	✓	✓
	Succinate dehydrogenase flavoprotein	sdhA	✓	✓	✓[n]	✓	✓	✓[n]	✓[n]	-	✓	✓
	Fumarate hydratase (class I)	fumA/B	-	✓	✓[n]	✓	✓	✓[n]	✓[n]	-	✓	✓
	Fumarate hydratase (class II)	fumC	✓	-	-	✓	✓	✓[n]	✓[n]	-	-	-
	Malate dehydrogenase (NAD+)	MDH	✓	✓	✓[n]	✓	✓	✓[n]	✓[n]	✓[?]	✓	✓[?]
	Malate quinone oxidoreductase	MQO	-	✓	✓[n]	✓	-	-	-	✓	✓	✓

Function	Protein	Abbr.	Organism									
			Ciliates	Dinozoa				Apicomplexa				
			Tet	*Pm*	*Po*	Dino	*Prom*	*Cv*	*Vb*	*Cp*	*Tg*	Plas/Pir
Electron transport chain	NADH dehydrogenase (complex I)	NDH1	✓	-	-	-	-	-	-	-	-	-
	Alternative NAD(P)H dehydrogenase (NDH)	NDH2	-	✓	✓[n]	✓	✓	✓[n]	✓[n]	✓	✓	✓
	Succinate dehydrogenase flavoprotein (CII)	SDHA	✓	✓	✓[n]	✓	✓	✓[n]	✓[n]	-	✓	✓
	Succinate dehydrogenase Fe-S (CII)	SDHB	✓	✓	✓[n]	✓	✓	✓[n]	✓[n]	-	✓	✓
	Ubiquinol cytochrome c reductase Rieske Fe-S (CIII)	qcrA	✓	✓	✓[n]	✓	✓	-	✓[n]	-	✓	✓
	Cytochrome c1 (CIII)	Cyt1	✓	✓	✓[n]	✓	✓	-	✓[n]	-	✓	✓
	Ubiquinol-cytochrome c reductase hinge (CIII)	UCR_hinge	-	✓	✓[n]	✓	✓	-	✓[n]	-	✓	✓
	Ubiquinol:cytochrome c oxidoreductase core (CIII)	pqqL	✓	✓	✓[n]	✓	✓	✓[n]	✓[n]	-	✓	✓
	Cytochrome b (CIII)	COB	✓	✓	✓	✓	✓	-	✓[F]	-	✓	✓
	Cytochrome c	CYC	✓	✓	✓[n?]	✓	✓	✓[n]	✓[n]	-	✓	✓
	Cytochrome oxidase I (CIV)	COXI	✓	✓	✓	✓	✓	✓	✓[F]	-	✓	✓
	Cytochrome oxidase II (CIV)	COX2	✓	✓	✓[n,a/b]	✓	✓	✓[n,a/b]	✓[n,a/b]	-	✓	✓
	Cytochrome oxidase III (CIV)	COX3	-	-	-	✓	✓	-	-	-	✓	✓
	Alternative oxidase	AOX	✓	✓	✓[n]	✓	✓	✓[n]	✓[n]	✓	-	-
Additional dehydrogenases	Dihydrooratate dehydrogenase	DHOD	✓	✓	✓[n]	✓	✓	✓[n]	✓[n]	-	✓	✓
	Glycerol 3-phosphate dehydrogenase	GlpA	✓	✓	✓[n]	-	✓	✓[n]	✓[n]	-	✓	✓
	Cytochrome b2	COB2	-	✓	✓[n]	✓	✓	✓[n]	-	-	-	-
ATP synthase complex	ATP synthase alpha chain (F1)	ATPA	✓	✓	✓[n]	✓	✓	✓[n]	✓[n]	✓	✓	✓
	ATP synthase beta chain (F1)	ATPB	✓	✓	✓[n]	✓	✓	✓[n]	✓[n]	✓	✓	✓
	ATP synthase gamma (F1)	ATPG	✓	✓	✓[n]	✓	✓	✓[n]	✓[n]	-	✓	✓
	ATP synthase delta (F1)	ATPD	-	✓	✓[n]	✓	✓	✓[n]	✓[n]	-	✓	✓
	ATP synthase epsilon (F1)	ATPE	-	✓	✓[n]	✓	✓	✓[n]	✓[n]	-	✓	✓
	ATP synthase lipid binding (F0)	ATPsyntLB	✓	✓	✓[n]	✓	✓	✓[n]	✓[n]	-	✓	✓
	Oligomycin sensitivity-conferring protein	Oscp	✓	✓	✓[n]	✓	✓	✓[n]	✓[n]	-	✓	✓
	a (F0)		-	-	-	-	-	-	-	-	-	✓
	b (F0)		-	-	-	-	-	-	-	-	-	✓

Function	Protein	Abbreviation	Organism									
			Ciliates	Dinozoa				Apicomplexa				
			Tet	*Pm*	*Po*	Dino	*Prom*	*Cv*	*Vb*	*Cp*	*Tg*	Plas/Pir
Mitochondrial carrier proteins	ADP/ATP translocase	ATPtrans	✓	✓	✓^{n}	✓	✓	✓^{n}	✓^{n}	✓	✓	✓
	Mitochondrial phosphate carrier	MPC	✓	✓	✓^{n}	✓	✓	✓^{n}	✓^{n}	✓	✓	✓
	Uncoupling mitochondrial protein	UMP	✓	-	✓^{n}	✓	✓	✓^{n}	✓^{n}	-	-	-
	Ketoglutarate/malate translocator	KMT	✓	✓	✓^{n}	✓	✓	✓^{n}	✓^{n}	✓	✓	✓
	Aspartate/glutamate carrier	AGC	✓$^{ID?}$	-	-	✓	-	-	-	-	-	-
	Oxodicarboxylate carrier	ODC	-	✓	✓^{n}	✓	✓	-	-	-	-	-
	Dicarboxylate carrier	DIC	-	✓	✓^{n}	✓	✓	-	-	-	-	-
	Tricarboxylate carrier	TRIC	✓	-	-	✓	✓	-	-	-	-	-
	Acyl carnitine carrier	ACC	✓	-	-	✓	✓	-	✓^{n}	-	-	-
Heme biosynthesis	Glutamyl-tRNA reductase	HemA	-	-	-	✓^{P}	✓	-	-	-	-	-
	Glutamate semialdehyde aminotransferase	HemL	-	-	-	✓^{P}	✓	-	-	-	-	-
	Delta-aminolaevulinic acid synthase	ALAS	✓	✓	✓^{n}	✓^{H}	-	✓^{n}	✓^{n}	-	✓	✓*
	Delta-aminolaevulinic acid dehydratase	HemB	✓	✓	✓^{n}	✓	✓	✓^{n}	✓^{n}	-	✓^{P}	✓^{P*}
	Porphobilinogen deaminase	HemC	✓	✓	✓^{n}	✓	✓	✓^{n}	✓^{n}	-	✓^{P}	✓^{P*}
	Uroporphyrinogen III synthase	HemD	-	-	-	-	-	-	-	-	✓^{P}	-
	Uroporphyrinogen III decarboxylase	HemE	✓	✓	✓^{n}	✓	✓	✓^{n}	✓^{n}	-	✓^{P}	✓^{P*}
	Uroporphyrinogen III oxidase	HemF	✓	-	-	✓	✓	✓^{n}	✓^{n}	-	✓	✓*
	Protoporphyrinogen oxidase	HemG	-	-	-	-	✓	✓^{n}	✓^{n}	-	✓	✓*
	Ferrochelatase	HemH	✓	-	-	✓	✓	✓^{n}	✓^{n}	-	✓	✓*
Fe-S cluster biosynthesis	Cysteine desulfurase	IscS	✓	✓	✓^{n}	✓	✓	✓^{n}	✓^{n}	✓	✓	✓
	Ferredoxin	mtFd	✓	✓	✓^{n}	✓	✓	✓^{n}	✓^{n}	✓	✓	✓
	Ferredoxin-NADP+ reductase	mtFNR	✓	✓	✓^{n}	-	✓	✓^{n}	✓^{n}	✓	✓	✓
	ISC scaffold protein	IscU	✓	✓	✓^{n}	✓	✓	✓^{n}	✓^{n}	✓	✓	✓
	Iron-Sulphur cluster assembly accessory protein	IscA	✓	✓	✓^{n}	✓	✓	✓^{n}	✓^{n}	-	✓	✓
	Adaptor and stabilizer of Nfs	Isd11	-	-	-	-	-	✓^{n}	✓^{n}	-	✓	✓
	Co-Chaperone	GrpE	✓	✓	✓^{n}	✓	✓	✓^{n}	✓^{n}	✓	✓	✓
	Frataxin	FXN	✓	✓	✓^{n}	✓	✓	✓^{n}	✓^{n}	✓	-	✓

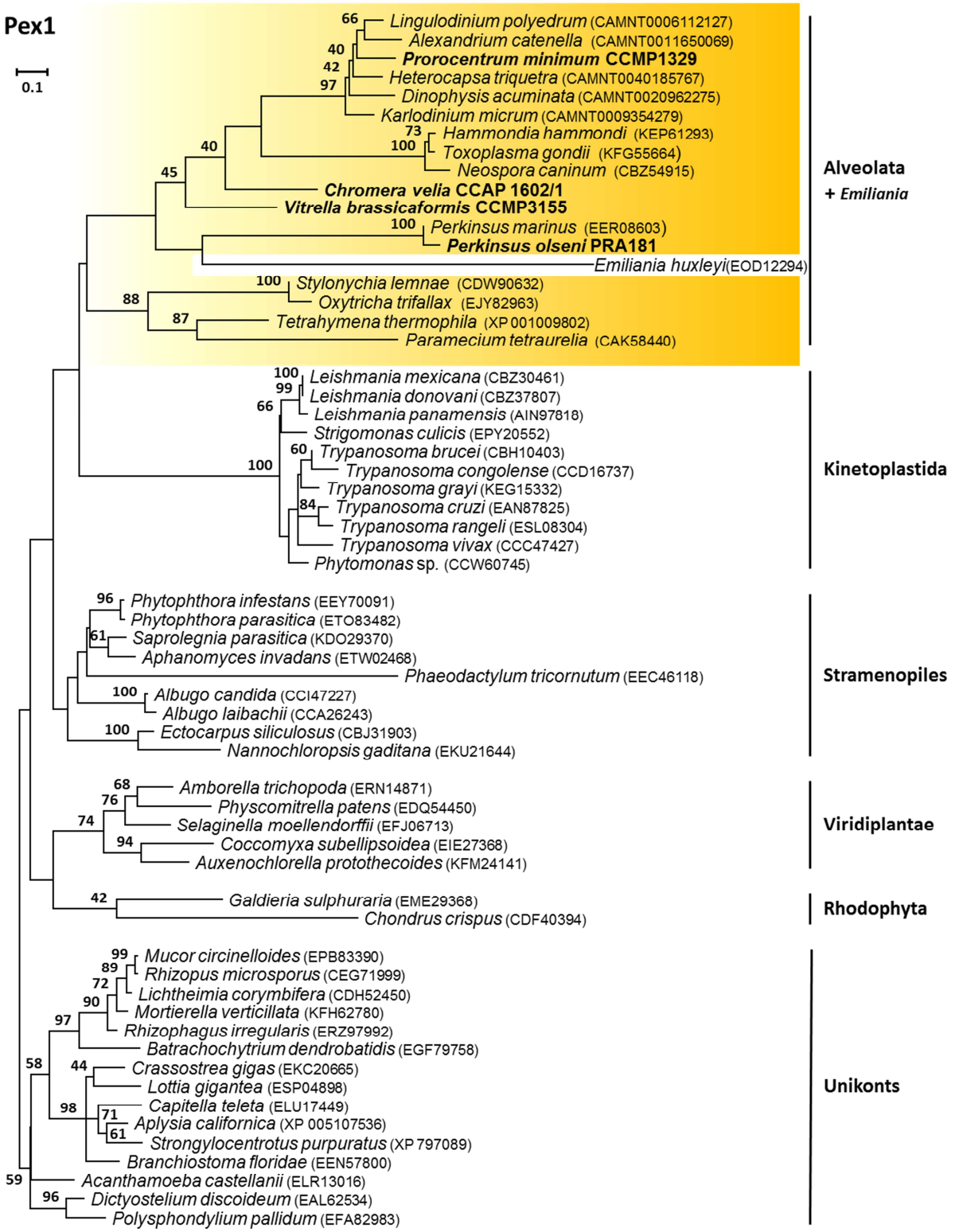


Figure S15: Phylogenetic maximum likelihood RAxML tree (LG+F+Γ4) based on 60 Pex1 sequences and 179 aa positions. Alveolate groups are marked by a yellow box. Alveolate sequences of our newly established transcriptomes are shown in bold and a larger font size.

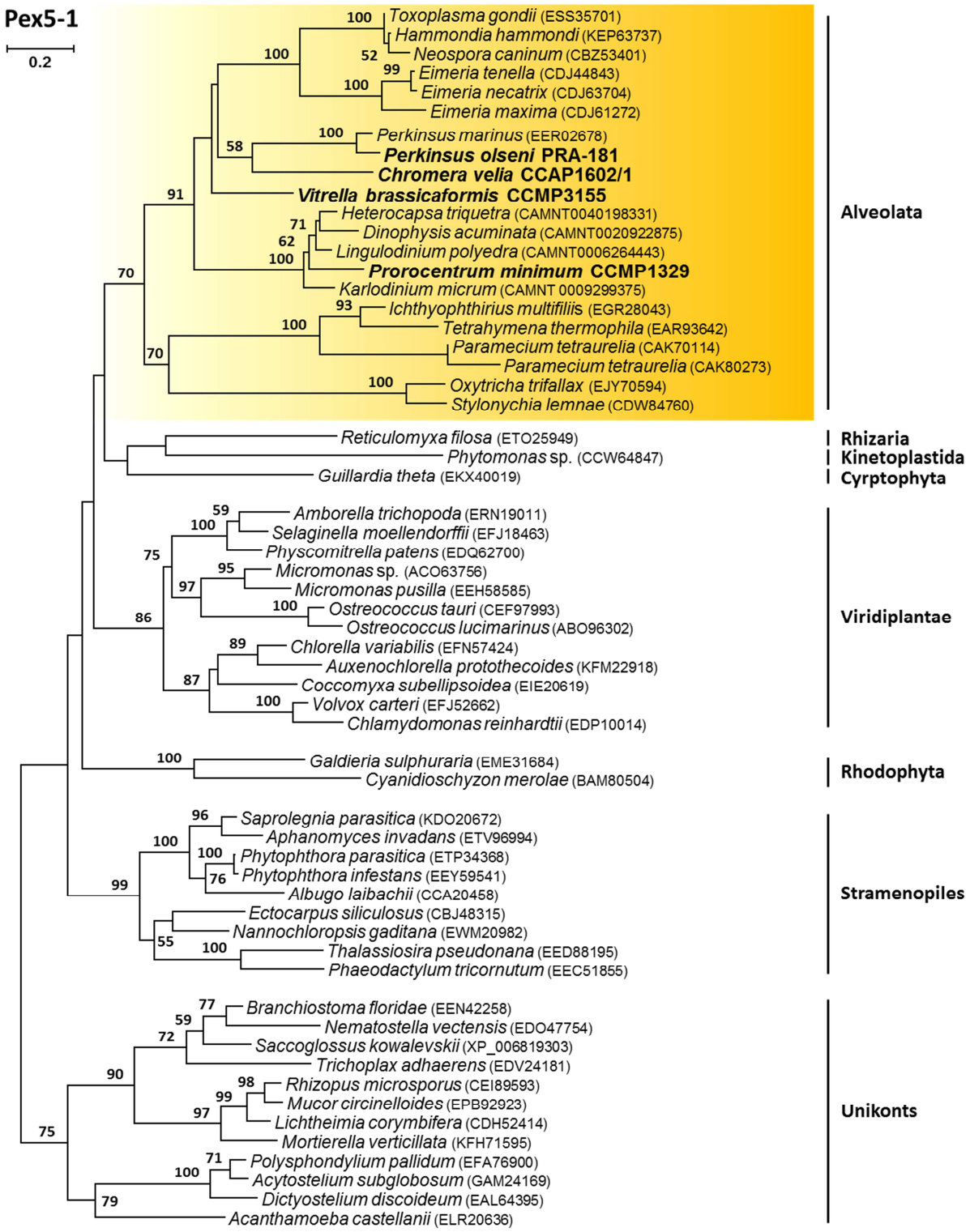


Figure S16: Phylogenetic maximum likelihood RAxML tree (LG+F+Γ4) based on 59 Pex5 sequences and 205 aa positions. Alveolate groups are marked by yellow box. Alveolate sequences of our newly established transcriptomes are shown in bold and a larger font size.

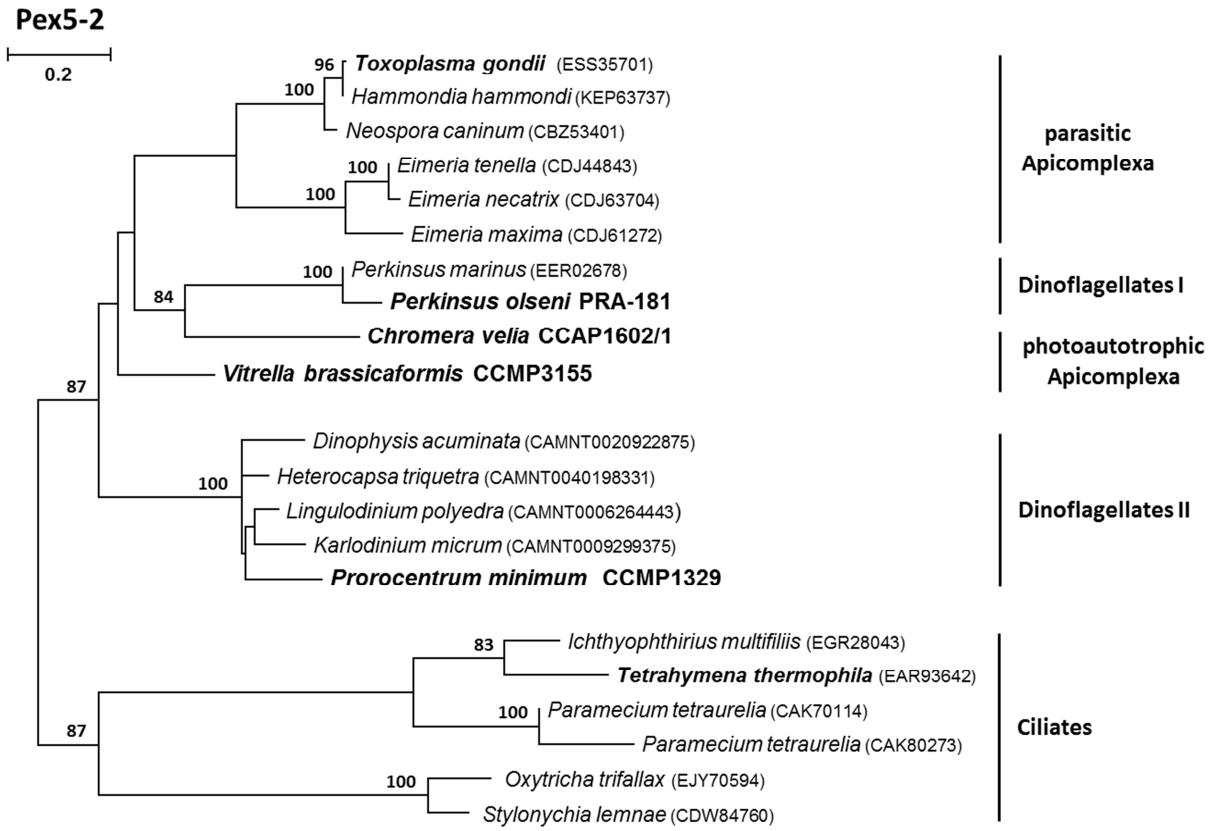


Figure S17: Phylogenetic maximum likelihood RAxML analysis (LG+F+Γ4) of alveolate Pex5 subtree based on 21 sequences and 209 aa positions. Alveolate sequences of our newly established transcriptomes are shown in bold and a larger font size.

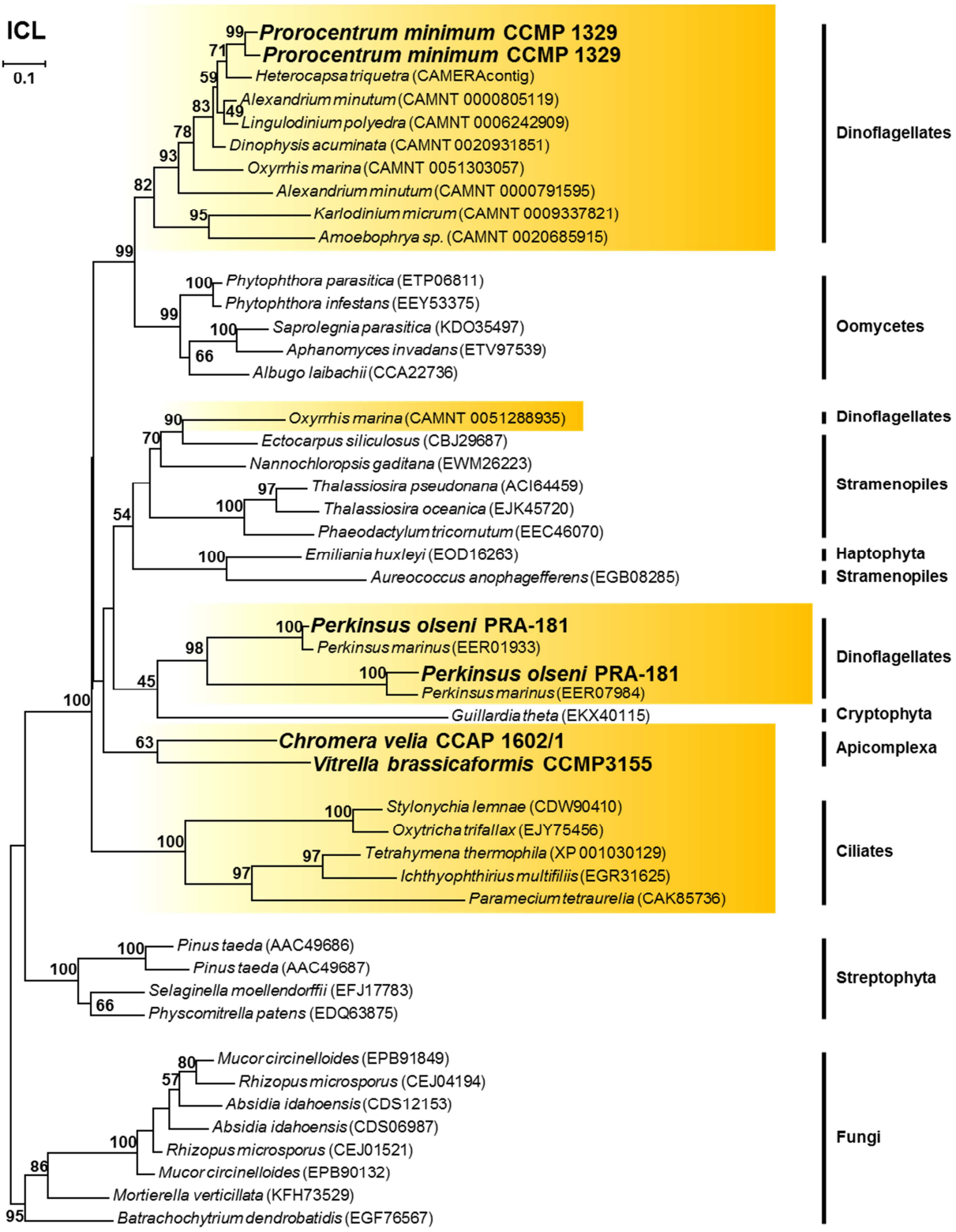


Figure S18: Phylogenetic maximum likelihood RAxML tree (LG+F+Γ4) based on 47 isocitrate lyase (ICL) sequences and 398 aa positions. Alveolate groups are marked by a yellow box. Alveolate sequences of our newly established transcriptomes are shown in bold and a larger font size.

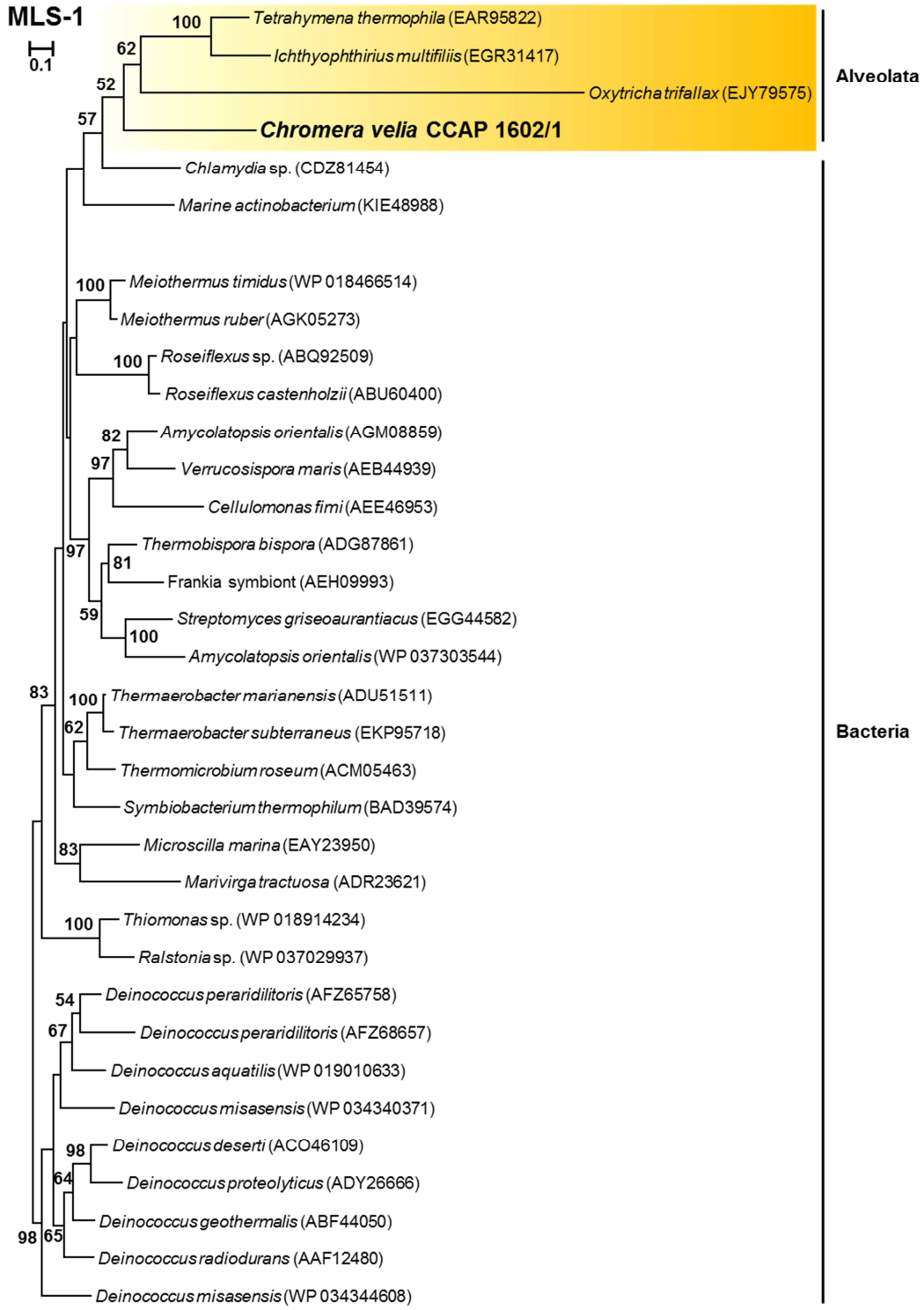


Figure S19: Phylogenetic maximum likelihood RAxML subanalysis (LG+F+Γ4) of the glyoxysomal malate synthase marker (MLS-1) based on 34 sequences and 415 aa positions. Alveolate groups are marked by a yellow box. Alveolate sequences of our newly established transcriptomes are shown in bold and a larger font size.

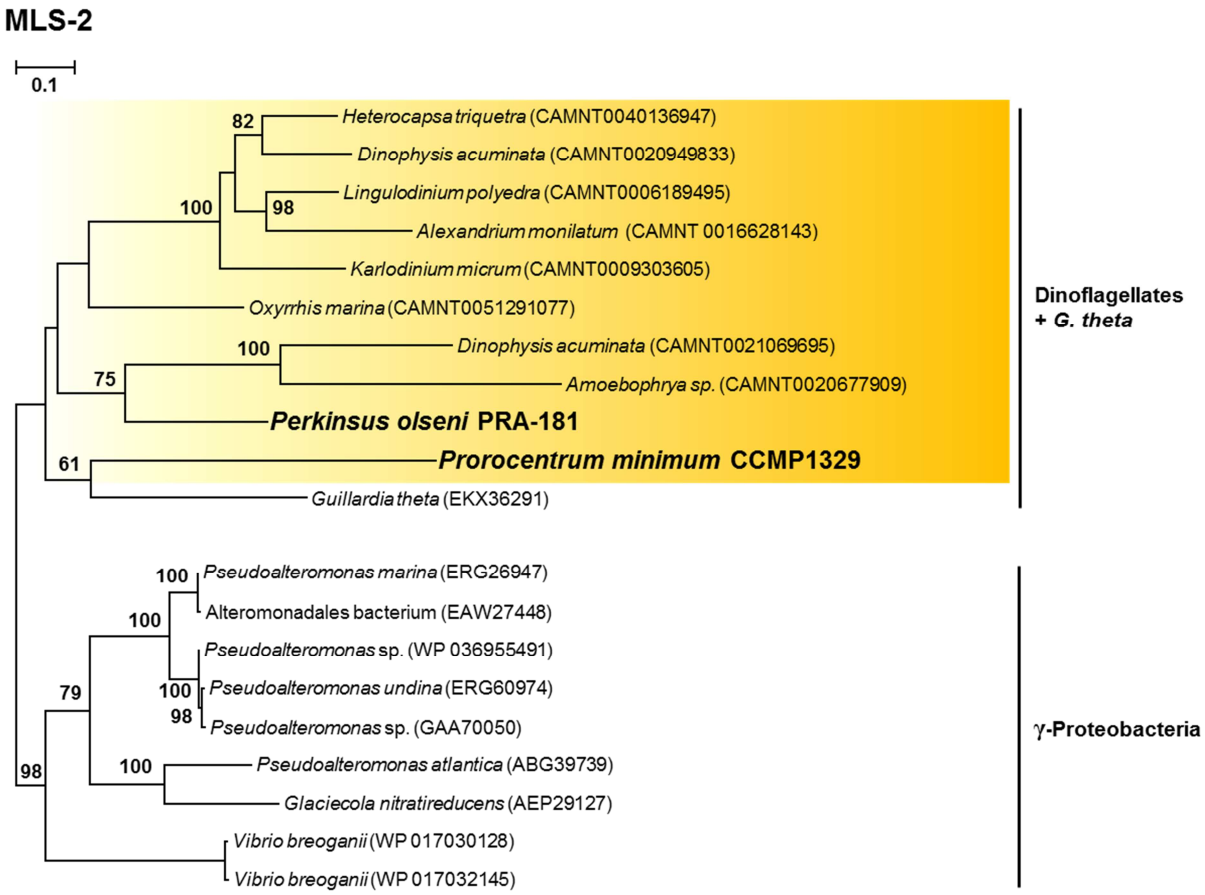


Figure S20: Phylogenetic maximum likelihood RAxML subanalysis (LG+F+Γ4) of the glyoxysomal malate synthase marker (MLS-2) based on 20 sequences and 513 aa positions. Alveolate groups are marked by a yellow box. Alveolate sequences of our newly established transcriptomes are shown in bold and a larger font size.

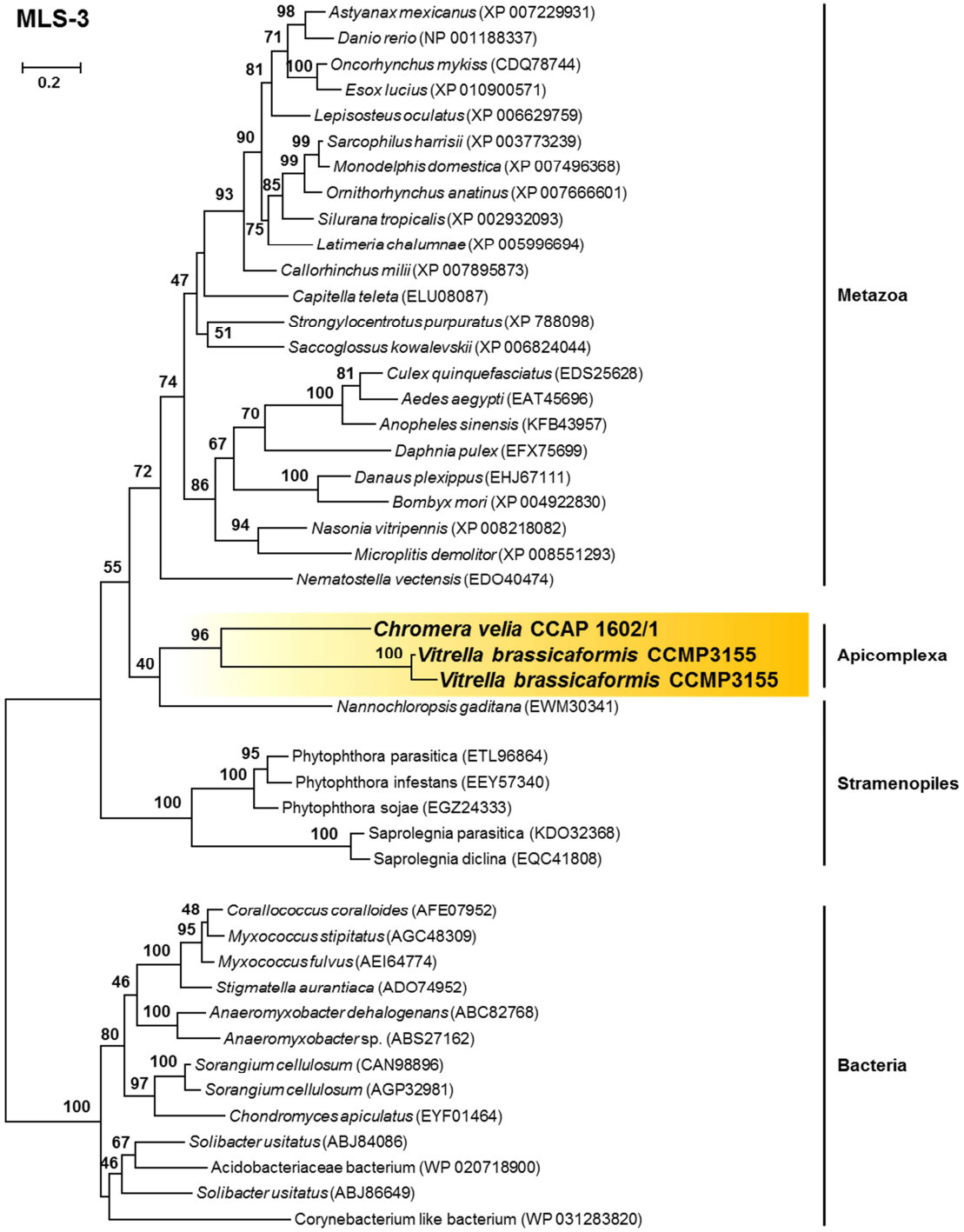


Figure S21: Phylogenetic maximum likelihood RAxML subanalysis (LG+F+Γ4) of the glyoxysomal malate synthase marker (MLS-3) based on 45 sequences and 319 aa positions. Alveolate groups are marked by a yellow box. Alveolate sequences of our newly established transcriptomes are shown in bold and a larger font size.

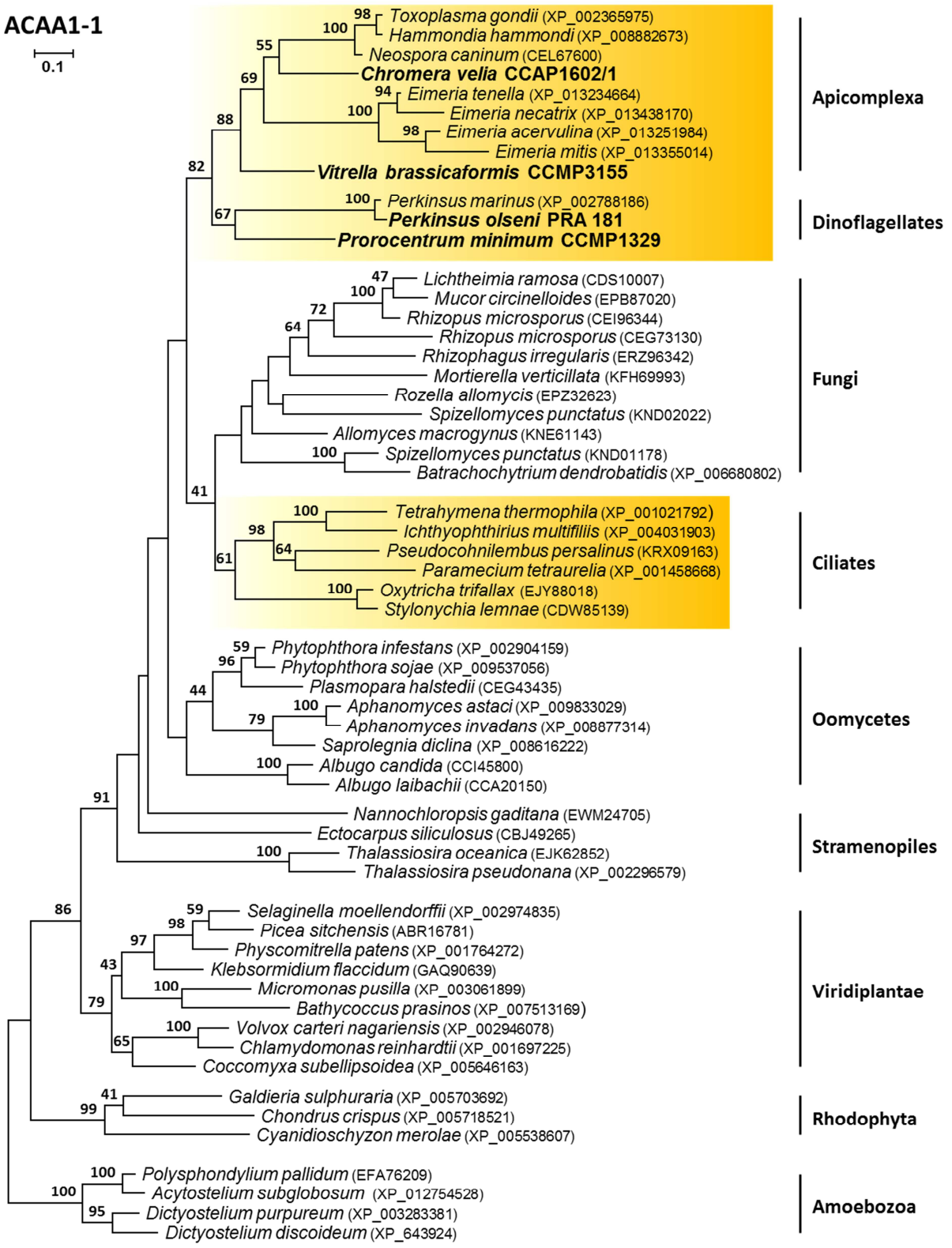


Figure S22: Phylogenetic maximum likelihood RAxML analysis (LG+F+Γ4) of the fatty acid oxidation marker ACAA1 based on 57 sequences and 326 aa positions. Alveolate groups are marked by a yellow box. Newly established myzozoan data are highlighted in bold and a larger font size.

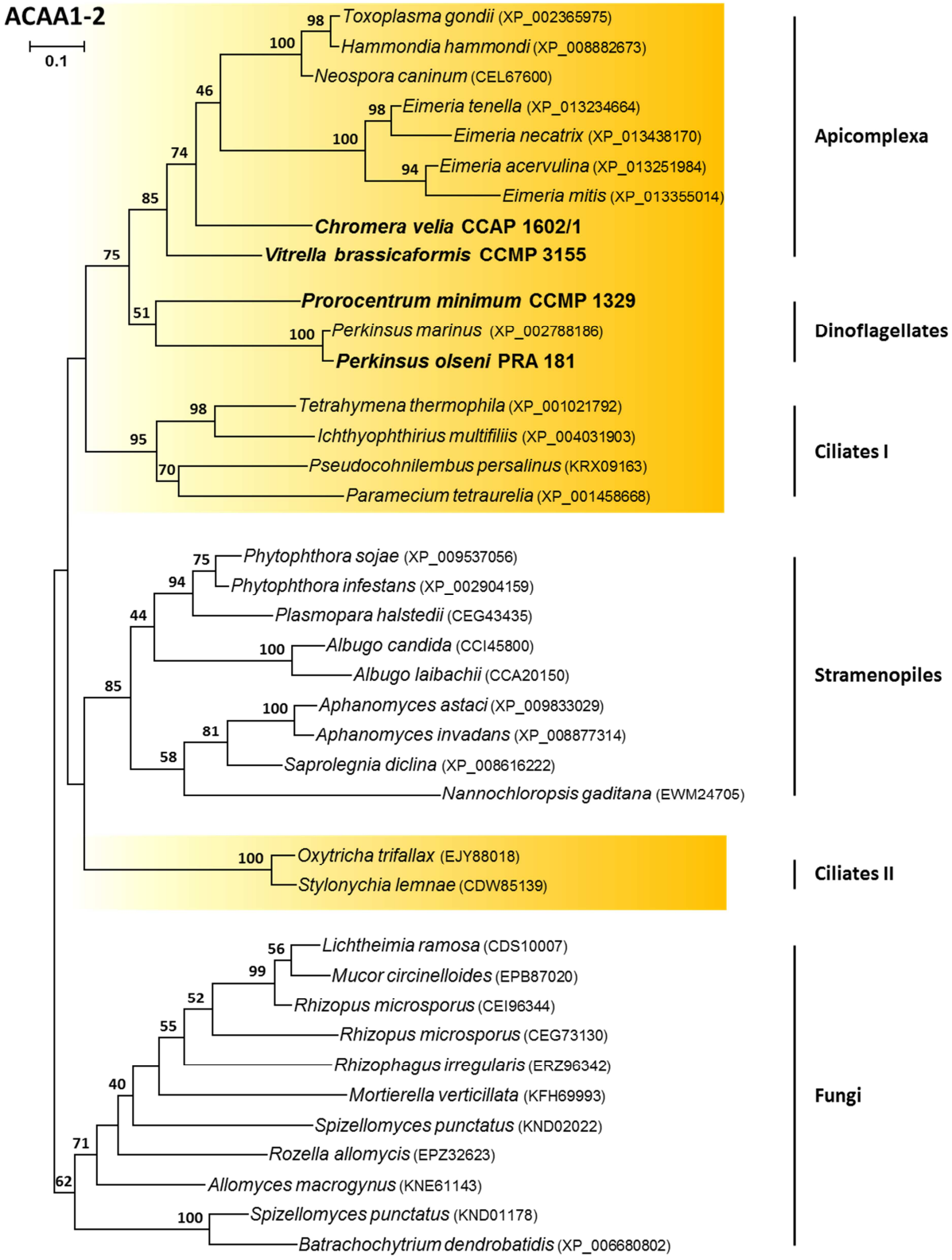


Figure S23: Phylogenetic maximum likelihood RAxML subanalysis (LG+F+Γ4) based on 38 ACAA1 sequences and 323 aa positions. Alveolate groups are marked by a yellow box. Newly established myzozoan data are highlighted in bold and a larger font size.

Table S8: Metabolic inventory of peroxisomal pathways in eight alveolate species. Presence of proteins are indicated by a check mark, absence by a dash. Subscripted characters contain the information about detected peroxisomal targeting sequences PTS1 and PTS2. * H_2O_2 releasing enzymes; [1] PTS1; [2] PTS2; [m] mitochondrial target sequence. Abbreviations for organisms: Tt, *Tetrahymena thermophila*; Po, *Perkinsus olseni;* Pm, *Prorocentrum minimum*; Vb, *Vitrella brassicaformis*; Cv, *Chromera velia*; Cp, *Cryptosporidium parvum*; Tg, *Toxoplasma gondii*; Pf, *Plasmodium falciparum.*

Functional category	Protein	Abbr.	Tt	Po	Pm	Vb	Cv	Cp	Tg	Pf
Fatty acid oxidation										
α-oxidation	2-hydroxyacyl-CoA lyase	HPCL2	✓	–	–	–	✓[1]	–	–	✓
	Phytanoyl-CoA hydrolase	PHYH	✓[2]	✓	✓[2]	✓[2]	✓[2]	–	–	–
β-oxidation	α-methylacyl-CoA-racemase	AMACR	✓	✓	–	–	–	–	–	–
	Acyl-CoA-oxidase	ACOX*	✓	–	✓[1]	✓[1]	✓[1]	–	✓[1]	–
	Multifunctional protein	DBP	✓[1]	✓[1]	✓[1]	✓[1]	✓[1]	–	✓[1]	–
	Sterole carrier protein 2	SCP-2	✓	✓[1]	✓	(✓)	✓[1]	–	✓	–
	Multifunctional protein	PBE	✓	✓[1]	✓[1]	–	–	–	–	–
	Acetyl-CoA acyltransferase 1	ACAA1	✓	✓[2]	✓[2]	✓[2]	✓[2]	–	✓	–
	2,4-dienoyl-CoA reductase	PDCR	✓	–	✓[1]	✓	✓[2]	–	✓[1]	–
	δ(3,5)-δ(2,4)-dienoyl-CoA isomerase	ECH	✓	✓	✓[1]	✓[1,2]	✓[1]	–	✓	–
	ATP-binding cassette, subfamily D	ABCD	✓	✓	✓	✓	✓	✓	✓	–
	Long-chain acyl-CoA synthethase	ACSL	✓	✓	✓	✓	✓	✓	✓	✓
	Solute carrier family 27, member 2	VLACS	–	✓	✓[1]	✓	✓	–	✓	–
Other oxidation	Acyl-CoA thioesterase 8	PTE	–	✓[1]	–	–	–	–	–	–
	Nucleoside disphosphate-linked m.	NUDT19	–	–	–	–	✓	–	–	–
Amino acid metabolism	Multifunctional protein	AGT	–	✓[1]	✓	✓	✓	–	✓	–
	D-amino-acid oxidase	DAO*	✓	–	–	–	–	–	–	–
	Isocitrate dehydrogenase	IDH	✓	✓	✓	✓	✓	–	✓[1]	✓[m]
	N1-acetylpolyamine oxidase	PAOX*	–	✓[1]	✓	✓	✓	–	–	–
	L-pipecolate oxidase	PIPOX*	–	✓	✓	✓[1]	✓[1]	–	–	–
	hydroxymethylgluatryl-CoA lyase	HMGCL	✓	–	–	✓	✓	–	✓	–
	(S)-2-hydroxy-acid oxidase	HAO*	✓	–	–	✓[1]	✓[1]	–	–	–
Antioxidant system										
Hydrogen peroxide metabolism	Catalase	CAT	✓[1]	–	–	✓	–	–	✓[1]	–
	Superoxide dismutase	SOD*	✓	✓	–	–	–	✓	✓	✓
	Nitric-oxide synthase, inducible	INOS	–	–	✓	–	–	–	–	–
	Peroxiredoxin 1	PRDX1	–	✓	–	✓	✓	✓	✓	✓[1]
	Peroxiredoxin 5	PRDX5	–	✓	✓	–	–	–	–	–
Glutathione metabolism	Glutathione S-transferase kappa 1	GSTK1	–	✓	✓[1]	✓[1]	✓[1]	–	–	–
Etherphospholipid biosynthesis	Dihydroxyacetone phosphate acyltr.	DHAPAT	✓[1,2]	✓[1]	✓	✓	✓	–	✓	–
	Alkyldihydroxyacetonephosphate syn.	AGPS	✓	–	✓[2]	✓[2]	✓	–	–	–
	Fatty acyl-CoA reductase	FAR	✓[1,2]	–	✓	✓	✓	–	✓	–
Purine metabolism	Xanthine dehydrogenase	XDH*	–	✓	✓[2]	✓	✓	–	–	–
Retinol metabolism	Dehydrogenase/reductase SDR family	DHRS4	✓	✓[1]	–	✓[1]	✓[1]	–	–	–
Sterol precursor biosynthesis	Mevalonate kinase	MVK	✓[2]	–	–	–	–	–	–	–
	Phosphomevalonate kinase	PMVK	–	–	✓	–	–	–	–	–

Table S9: Peroxisomal biogenesis markers and PMPs in other whole-genome sequenced Apicomplexa via blastp and tblastn analyses in NCBI.

Species	Group	Peroxins														PMP		
		1	2	3	4	5	6	7	10	11	12	14	16	19	22	22	34	70
P. falciparum	Aconoidasida	-	-	-	✓	-	-	-	-	-	-	-	-	-	✓	-	-	-
P. vivax	Aconoidasida	-	-	-	✓	-	-	-	-	-	-	-	-	-	✓	-	-	-
P. yoelii	Aconoidasida	-	-	-	✓	-	-	-	-	-	-	-	-	-	✓	-	-	-
B. bovis	Aconoidasida	-	-	-	✓	-	-	-	-	-	-	-	-	-	✓	-	-	-
B. bigemina	Aconoidasida	-	-	-	✓	-	-	-	-	-	-	-	-	-	✓	-	-	-
T. annulata	Aconoidasida	-	-	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-
T. parva	Aconoidasida	-	-	-	✓	-	-	-	-	-	-	-	-	-	✓	-	-	-
T. gondii	Conoidasida	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	-	✓	✓
H. hammondi	Conoidasida	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	-	✓	-	✓	✓
E. tenella	Conoidasida	✓	✓	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	-	-	✓	-
E. praecox	Conoidasida	✓	✓	-	✓	✓	-	✓	✓	✓	✓	✓	✓	-	-	-	✓	-
N. caninum	Conoidasida	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	-	✓	✓
S. neurona	Conoidasida	-	✓	✓	✓	✓	-	✓	✓	✓	✓	-	-	-	✓	-	✓	✓
C. cayetanensis	Conoidasida	-	✓	-	✓	✓	-	✓	✓	✓	✓	✓	✓	-	-	-	✓	-
C. parvum	Conoidasida	-	-	-	✓	-	-	-	-	-	-	-	-	-	✓	-	-	-
C. hominis	Conoidasida	-	-	-	✓	-	-	-	-	-	-	-	-	-	✓	-	-	-
G. niphandrodes	Conoidasida	-	-	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-
A. taiwanensis	Conoidasida	✓	-	-	-	✓	-	✓	-	-	-	-	-	-	-	-	✓	✓
C. velia	Chromerida	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	✓	✓
V. brassicaformis	Chromerida	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table S10: Important peroxisomal metabolic markers in other whole-genome sequenced Apicomplexa via blastp and tblastn analyses in NCBI. Abbreviations due to the text.

Species	Group	GC		AoxS	FAO				EPL
		ICL	MLS	CAT	ACAA1	ACOX	ECH	DBP	DHAPAT
P. falciparum	Aconoidasida	-	-	-	-	-	-	-	-
P. vivax	Aconoidasida	-	-	-	-	-	-	-	-
P. yoelii	Aconoidasida	-	-	-	-	-	-	-	-
B. bovis	Aconoidasida	-	-	-	-	-	-	-	-
B. bigemina	Aconoidasida	-	-	-	-	-	-	-	-
T. annulata	Aconoidasida	-	-	-	-	-	-	-	-
T. parva	Aconoidasida	-	-	-	-	-	-	-	-
T. gondii	Conoidasida	-	-	✓	✓	✓	✓	✓	✓
H. hammondi	Conoidasida	-	-	✓	✓	✓	✓	✓	✓
E. tenella	Conoidasida	-	-	✓	✓	✓	✓	✓	✓
E. praecox	Conoidasida	-	-	✓	✓	✓	✓	✓	✓
N. caninum	Conoidasida	-	-	✓	✓	✓	✓	✓	✓
S. neurona	Conoidasida	-	-	✓	✓	✓	✓	✓	✓
C. cayetanensis	Conoidasida	-	-	✓	✓	✓	✓	✓	✓
C. parvum	Conoidasida	-	-	-	-	-	-	-	-
C. hominis	Conoidasida	-	-	-	-	-	-	-	-
G. niphandrodes	Conoidasida	-	-	✓	-	-	-	-	-
A. taiwanensis	Conoidasida	-	✓	✓	✓	✓	-	✓	-
C. velia	Chromerida	✓	✓	-	✓	✓	✓	✓	✓
V. brassicaformis	Chromerida	✓	✓	✓	✓	✓	✓	✓	✓

Table S11: Peroxisomal biogenesis markers and PMPs in other sequenced dinoflagellate species and the ciliate reference *Tetrahymena thermophila* inferred via blastp and tblastn searches in NCBI.

Species	Group	Peroxins														PMP		
		1	2	3	4	5	6	7	10	11	12	14	16	19	22	22	34	70
P. minimum	Prorocentrales	✓	✓	✓	✓	✓	✓	✓	-	✓	-	-	-	✓	✓	-	✓	✓
S. minutum	Suessiales	✓	-	-	-	✓	-	-	-	-	✓	-	-	-	-	-	-	✓
P. marinus	Perkinsea	✓	-	✓	✓	✓	✓	✓	-	✓	✓	✓	✓	✓	✓	✓	✓	✓
P. olseni	Perkinsea	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	✓	✓	✓	✓	✓	✓	✓
T. thermophila	Ciliata	✓	✓	-	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	-	-	✓	✓

Table S12: Important peroxisomal metabolic markers in other sequenced dinoflagellate species and the ciliate reference via blastp and tblastn analyses in NCBI. Abbreviations due to the text.

Species	Group	GC		AoxS	FAO				EPL
		ICL	MLS	CAT	ACAA1	ACOX	ECH	DBP	DHAPAT
P. minimum	Prorocentrales	✓	✓	-	✓	✓	✓	✓	✓
S. minutum	Suessiales	✓	✓	-	✓	✓	✓	✓	-
P. marinus	Perkinsea	✓	✓	-	✓	✓	✓	✓	✓
P. olseni	Perkinsea	✓	✓	-	✓	-	✓	✓	✓
T. thermophila	Ciliata	✓	✓	✓	✓	✓	✓	✓	✓

6. References

Abrahamsen, M. S.; Templeton, T. J.; Enomoto, S.; Abrahante, J. E.; Zhu, G.; Lancto, C. A. et al. (2004): Complete genome sequence of the apicomplexan, Cryptosporidium parvum. In *Science (New York, N.Y.)* 304 (5669), pp. 441–445.

Adams, K. L.; Palmer, J. D. (2003): Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. In *Molecular Phylogenetics and Evolution* 29 (3), pp. 380–395.

Adl, S. M.; Simpson, A. G. B.; Lane, C. E.; Lukes, J.; Bass, D.; Bowser, S. S. et al. (2012): The revised classification of eukaryotes. In *The Journal of eukaryotic microbiology* 59 (5), pp. 429–493.

Adl, S. M.; Leander, B. S.; Simpson, A. G. B.; Archibald, J. M.; Anderson, O. Roger; Bass, D. et al. (2007): Diversity, nomenclature, and taxonomy of protists. In *Systematic biology* 56 (4), pp. 684–689.

Adl, S. M.; Simpson, A. G. B.; Farmer, M. A.; Andersen, R. A.; Anderson, O. R.; Barta, J. R. et al. (2005): The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. In *The Journal of eukaryotic microbiology* 52 (5), pp. 399–451.

Akhmanova, A.; Voncken, F.; van Alen, T.; van Hoek, A.; Boxma, B.; Vogels, G. et al. (1998): A hydrogenosome with a genome. In *Nature* 396 (6711), pp. 527–528.

Ali, A. B.; Baere, R. de; van der Auwera, G.; Wachter, R. de; van de Peer, Y. (2001): Phylogenetic relationships among algae based on complete large-subunit rRNA sequences. In *International journal of systematic and evolutionary microbiology* 51 (3), pp. 737–749.

Allen, J. F. (2003): The function of genomes in bioenergetic organelles. In *Philosophical Transactions of the Royal Society B: Biological Sciences* 358 (1429), pp. 19–38.

Allen, C. A.; Hakansson, G.; Allen, J. F. (1995): Redox conditions specify the proteins synthesised by isolated chloroplasts and mitochondria. In *Redox report : communications in free radical research* 1 (2), pp. 119–123.

Allen, J. F. (1993): Control of gene expression by redox potential and the requirement for chloroplast and mitochondrial genomes. In *Journal of theoretical biology* 165 (4), pp. 609–631.

Altermann, W.; Kazmierczak, J.; Oren, A.; Wright, D. T. (2006): Cyanobacterial calcification and its rock-building potential during 3.5 billion years of Earth history. In *Geobiology* 4 (3), pp. 147–166.

Antonenkov, V. D.; Hiltunen, J. K. (2012): Transfer of metabolites across the peroxisomal membrane. In *Biochimica et biophysica acta* 1822 (9), pp. 1374–1386.

Archibald, J. M.; Lane, C. E. (2009): Going, going, not quite gone: nucleomorphs as a case study in nuclear genome reduction. In *The Journal of heredity* 100 (5), pp. 582–590.

Archibald, J. M. (2008): Plastid evolution: remnant algal genes in ciliates. In *Current Biology* 18 (15), R663-R665.

Archibald, J. M.; Rogers, M. B.; Toop, M.; Ishida, K.-I.; Keeling, P. J. (2003): Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga Bigelowiella natans. In *Proceedings of the National Academy of Sciences of the United States of America* 100 (13), pp. 7678–7683.

Arisue, N.; Hashimoto, T.; Mitsui, H.; Palacpac, N. M.; Kaneko, A.; Kawai, S. et al. (2012): The Plasmodium apicoplast genome: conserved structure and close relationship of P. ovale to rodent malaria parasites. In *Molecular biology and evolution* 29 (9), pp. 2095–2099.

Armbrust, E. V.; Berges, J. A.; Bowler, C.; Green, B. R.; Martinez, D.; Putnam, N. H. et al. (2004): The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism. In *Science* 306 (5693), pp. 79–86.

Aronesty, E. (2011): Command-line tools for processing biological sequencing data. ea-utils. Expression Analysis. Durham, NC. Available online at http://code.google.com/p/ea-utils.

Ascaso, C.; Wierzchos, J.; Speranza, M.; Gutiérrez, J. C.; González, A. M.; los Ríos, A. de; Alonso, J. (2005): Fossil protists and fungi in amber and rock substrates. In *Micropaleontology* 51 (1), pp. 59–72.

Ashley, E. A.; Dhorda, M.; Fairhurst, R. M.; Amaratunga, C.; Lim, P.; Suon, S. et al. (2014): Spread of artemisinin resistance in Plasmodium falciparum malaria. In *The New England journal of medicine* 371 (5), pp. 411–423.

Aurrecoechea, C.; Brestelli, J.; Brunk, B. P.; Dommer, J.; Fischer, S.; Gajria, B. et al. (2009): PlasmoDB: a functional genomic database for malaria parasites. In *Nucleic acids research* 37 (Database issue), D539-43.

Bachvaroff, T. R.; Gornik, S. G.; Concepcion, G. T.; Waller, R. F.; Mendez, G. S.; Lippmeier, J. C.; Delwiche, C. F. (2014): Dinoflagellate phylogeny revisited: using ribosomal proteins to resolve deep branching dinoflagellate clades. In *Molecular Phylogenetics and Evolution* 70, pp. 314–322.

Bachvaroff, T. R.; Handy, S. M.; Place, A. R.; Delwiche, C. F. (2011): Alveolate phylogeny inferred using concatenated ribosomal proteins. In *The Journal of eukaryotic microbiology* 58 (3), pp. 223–233..

Bachvaroff, T. R.; Place, A. R.; Coats, D. W. (2009): Expressed sequence tags from Amoebophrya sp. infecting Karlodinium veneficum: comparing host and parasite sequences. In *The Journal of eukaryotic microbiology* 56 (6), pp. 531–541.

Bachvaroff, T. R.; Place, A. R. (2008): From stop to start: tandem gene arrangement, copy number and trans-splicing sites in the dinoflagellate Amphidinium carterae. In *PloS one* 3 (8), e2929.

Bachvaroff, T. R.; Sanchez Puerta, M. V.; Delwiche, C. F. (2005): Chlorophyll c-containing plastid relationships based on analyses of a multigene data set with all four chromalveolate lineages. In *Molecular biology and evolution* 22 (9), pp. 1772–1782.

Bahl, A.; Davis, P. H.; Behnke, M.; Dzierszinski, F.; Jagalur, M.; Chen, F. et al. (2010): A novel multifunctional oligonucleotide microarray for Toxoplasma gondii. In *BMC genomics* 11, p. 603.

Balaban, R. S.; Nemoto, S.; Finkel, T. (2005): Mitochondria, oxidants, and aging. In *Cell* 120 (4), pp. 483–495.

Balabaskaran, N. P.; Dudkina, N. V.; La Kane; van Eyk, J. E.; Boekema, E. J.; Mather, M. W.; Vaidya, A. B. (2010): Highly divergent mitochondrial ATP synthase complexes in Tetrahymena thermophila. In *PLoS biology* 8 (7), e1000418.

Bannai, H.; Tamada, Y.; Maruyama, O.; Nakai, K.; Miyano, S. (2002): Extensive feature detection of N-terminal protein sorting signals. In *Bioinformatics (Oxford, England)* 18 (2), pp. 298–305.

Barbrook, A. C.; Santucci, N.; Plenderleith, L. J.; Hiller, R. G.; Howe, C. J. (2006): Comparative analysis of dinoflagellate chloroplast genomes reveals rRNA and tRNA genes. In *BMC genomics* 7, p. 297.

Barta, J. R.; Thompson, R. C. A. (2006): What is Cryptosporidium? Reappraising its biology and phylogenetic affinities. In *Trends in parasitology* 22 (10), pp. 463–468.

Baudhuin, P.; Müller, M.; Poole, B.; Duve, C. de (1965): Non-mitochondrial oxidizing particles (microbodies) in rat liver and kidney and in Tetrahymena pyriformis. In *Biochemical and biophysical research communications* 20 (1), pp. 53–59.

Baurain, D.; Brinkmann, H.; Petersen, J.; Rodriguez-Ezpeleta, N.; Stechmann, A.; Demoulin, V. et al. (2010): Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. In *Molecular biology and evolution* 27 (7), pp. 1698–1709.

Becker, B.; Hoef-Emden, K.; Melkonian, M. (2008): Chlamydial genes shed light on the evolution of photoautotrophic eukaryotes. In *BMC evolutionary biology* 8, p. 203.

Bender, A.; Krishnan, K. J.; Morris, C. M.; Taylor, G. A.; Reeve, A. K.; Perry, R. H. et al. (2006): High levels of mitochondrial DNA deletions in substantia nigra neurons in aging and Parkinson disease. In *Nature genetics* 38 (5), pp. 515–517.

Bendtsen, J. D.; Nielsen, H.; Heijne, G. von; Brunak, S. (2004): Improved prediction of signal peptides: SignalP 3.0. In *Journal of molecular biology* 340 (4), pp. 783–795.

Bessoff, K.; Sateriale, A.; Lee, K. K.; Huston, C. D. (2013): Drug repurposing screen reveals FDA-approved inhibitors of human HMG-CoA reductase and isoprenoid synthesis that block Cryptosporidium parvum growth. In *Antimicrobial agents and chemotherapy* 57 (4), pp. 1804–1814.

Bibby, B. T.; Dodge, J. D. (1973): The ultrastructure and cytochemistry of microbodies in dinoflagellates. In *Planta* 112 (1), pp. 7–16.

Blanchard, J. L.; Lynch, M. (2000): Organellar genes: why do they end up in the nucleus? In *Trends in genetics : TIG* 16 (7), pp. 315–320.

Blankenship, R. E. (2010): Early evolution of photosynthesis. In *Plant physiology* 154 (2), pp. 434–438.

Blum, J. J. (1982): The metabolic role of peroxisomes in Tetrahymena. In *Annals of the New York Academy of Sciences* 386, pp. 217–227.

Bodył, A. (2005): Do plastid-related characters support the chromalveolate hypothesis? In *Journal of Phycology* 41 (3), pp. 712–719.

Boetzer, M.; Henkel, C. V.; Jansen, H. J.; Butler, D.; Pirovano, W. (2011): Scaffolding pre-assembled contigs using SSPACE. In *Bioinformatics (Oxford, England)* 27 (4), pp. 578–579.

Bolte, K.; Gruenheit, N.; Felsner, G.; Sommer, M. S.; Maier, U.-G.; Hempel, F. (2011): Making new out of old: recycling and modification of an ancient protein translocation system during eukaryotic evolution. Mechanistic comparison and phylogenetic analysis of ERAD, SELMA and the peroxisomal importomer. In *BioEssays : news and reviews in molecular, cellular and developmental biology* 33 (5), pp. 368–376.

Bosch, S. S.; Kronenberger, T.; Meissner, K. A.; Zimbres, F. M.; Stegehake, D.; Izui, N. M. et al. (2015): Oxidative stress control by apicomplexan parasites. In *BioMed research international* 2015, p. 351289.

Boyd, E. S.; Thomas, K. M.; Dai, Y.; Boyd, J. M.; Outten, F. W. (2014): Interplay between oxygen and Fe-S cluster biogenesis: insights from the Suf pathway. In *Biochemistry* 53 (37), pp. 5834–5847.

Boyer, P. D. (1997): The ATP synthase--a splendid molecular machine. In *Annual review of biochemistry* 66, pp. 717–749.

Brunk, C. F. (2003): Complete sequence of the mitochondrial genome of Tetrahymena thermophila and comparative methods for identifying highly divergent genes. In *Nucleic acids research* 31 (6), pp. 1673–1682.

Bui, E. T.; Bradley, P. J.; Johnson, P. J. (1996): A common evolutionary origin for mitochondria and hydrogenosomes. In *Proceedings of the National Academy of Sciences* 93 (18), pp. 9651–9656.

Buick, R. (2010): Early life: Ancient acritarchs. In *Nature* 463 (7283), pp. 885–886.

Buick, R. (2008): When did oxygenic photosynthesis evolve? In *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 363 (1504), pp. 2731–2743.

Bullerwell, E. C. (Ed.) (2012): Organelle Genetics: Evolution of Organelle Genomes and Gene Expression. Berlin, Heidelberg: Springer Berlin Heidelberg.

Burger, G.; Zhu, Y.; Littlejohn, T. G.; Greenwood, S. J.; Schnare, M. N.; Lang, B. F.; Gray, M. W. (2000): Complete sequence of the mitochondrial genome of Tetrahymena pyriformis and comparison with Paramecium aurelia mitochondrial DNA. In *Journal of molecular biology* 297 (2), pp. 365–380.

Burki, F.; Imanian, B.; Hehenberger, E.; Hirakawa, Y.; Maruyama, S.; Keeling, P. J. (2014): Endosymbiotic gene transfer in tertiary plastid-containing dinoflagellates. In *Eukaryotic cell* 13 (2), pp. 246–255.

Burki, F.; Flegontov, P.; Obornik, M.; Cihlar, J.; Pain, A.; Lukes, J.; Keeling, P. J. (2012): Re-evaluating the green versus red signal in eukaryotes with secondary plastid of red algal origin. In *Genome biology and evolution* 4 (6), pp. 626–635.

Burki, F.; Hirakawa, Y.; Keeling, P. J. (2012): Intragenomic spread of plastid-targeting presequences in the coccolithophore Emiliania huxleyi. In *Molecular biology and evolution* 29 (9), pp. 2109–2112.

Burki, F.; Kudryavtsev, A.; Matz, M. V.; Aglyamova, G. V.; Bulman, S.; Fiers, M. et al. (2010): Evolution of Rhizaria: new insights from phylogenomic analysis of uncultivated protists. In *BMC evolutionary biology* 10, p. 377.

Burki, F.; Inagaki, Y.; Brate, J.; Archibald, J. M.; Keeling, P. J.; Cavalier-Smith, T. et al. (2009): Large-scale phylogenomic analyses reveal that two enigmatic protist lineages, telonemia and centroheliozoa, are related to photosynthetic chromalveolates. In *Genome biology and evolution* 1, pp. 231–238.

Burki, F.; Shalchian-Tabrizi, K.; Pawlowski, J. (2008): Phylogenomics reveals a new 'megagroup' including most photosynthetic eukaryotes. In *Biology letters* 4 (4), pp. 366–369.

Burreson, E. M.; Reece, K. S.; Dungan, C. F, (2005): Molecular, morphological, and experimental evidence support the synonymy of Perkinsus chesapeaki and Perkinsus andrewsi. In *The Journal of eukaryotic microbiology* 52 (3), pp. 258–270.

Butterfield, E. R.; Howe, C. J.; Nisbet, R. E. R. (2013): An analysis of dinoflagellate metabolism using EST data. In *Protist* 164 (2), pp. 218–236.

Butterfield, N. J. (2000): Bangiomorpha pubescens n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. In *Paleobiology* 26 (3), pp. 386–404.

Calandra, F. (1964): Sur un presume dinoflagelle Arpylorus nov gen du gothlandien de Tunisie. In *Comptes Rendus Hebdomadaires des Seances de l'Academie des Sciences* 258 (16), pp. 4112–4114.

Calhoun, M. W.; Thomas, J. W.; Gennis, R. B. (1994): The cytochrome oxidase superfamily of redox-driven proton pumps. In *Trends in biochemical sciences* 19 (8), pp. 325–330.

Camacho, C. (2016): BLAST+ Release Notes: National Center for Biotechnology Information (US).

Camoes, F.; Bonekamp, N. A.; Delille, H. K.; Schrader, M. (2009): Organelle dynamics and dysfunction: A closer link between peroxisomes and mitochondria. In *Journal of inherited metabolic disease* 32 (2), pp. 163–180.

Carreno, R. A.; Matrin, D. S.; Barta, J. R. (1999): Cryptosporidium is more closely related to the gregarines than to coccidia as shown by phylogenetic analysis of apicomplexan parasites inferred using small-subunit ribosomal RNA gene sequences. In *Parasitology Research* 85 (11), pp. 899–904.

Cavalier-Smith, T.; Chao, E. E. (2004): Protalveolate phylogeny and systematics and the origins of Sporozoa and dinoflagellates (phylum Myzozoa nom. nov.). In *European Journal of Protistology* 40 (3), pp. 185–212.

Cavalier-Smith, T. (2000): Membrane heredity and early chloroplast evolution. In *Trends in Plant Science* 5 (4), pp. 174–182.

Cavalier-Smith, T. (1999): Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. In *Journal of Eukaryotic Microbiology* 46 (4), pp. 347–366.

Cavalier-Smith, T. (1993): Kingdom protozoa and its 18 phyla. In *Microbiological reviews* 57 (4), pp. 953–994.

Chaban, Y.; Boekema, E. J.; Dudkina, N. V. (2014): Structures of mitochondrial oxidative phosphorylation supercomplexes and mechanisms for their stabilisation. In *Biochimica et biophysica acta* 1837 (4), pp. 418–426.

Chan, C. X.; Bhattacharya, D.; Reyes-Prieto, A. (2012): Endosymbiotic and horizontal gene transfer in microbial eukaryotes: Impacts on cell evolution and the tree of life. In *Mobile genetic elements* 2 (2), pp. 101–105.

Chan, D. C. (2006): Mitochondria: dynamic organelles in disease, aging, and development. In *Cell* 125 (7), pp. 1241–1252.

Charan, M.; Singh, N.; Kumar, B.; Srivastava, K.; Siddiqi, M. I.; Habib, S. (2014): Sulfur mobilization for Fe-S cluster assembly by the essential SUF pathway in the Plasmodium falciparum apicoplast and its inhibition. In *Antimicrobial agents and chemotherapy* 58 (6), pp. 3389–3398.

Chen, X.; Ma, H.-G.; Al-Rasheid, K. A. S.; Miao, M. (2015): Molecular data suggests the ciliate Mesodinium (Protista: Ciliophora) might represent an undescribed taxon at class level. In *Zoological Systematics* 40 (1), pp. 31–40.

Chesnick, J. M.; Kooistra, W. H. C.; Wellbrock, U.; Medlin, L. K. (1997): Ribosomal RNA Analysis Indicates a Benthic Pennate Diatom Ancestry for the Endosymbionts of the Dinoflagellates Peridinium foliaceum and Peridinium balticum (Pyrrhophyta). In *J Eukaryotic Microbiology* 44 (4), pp. 314–320.

Chesnick, J. M.; Morden, C. W.; Schmieg, A. M. (1996): Identity of the endosymbiont of Peridinium foliaceum (Pyrrophyta): analysis of the rbcLS operon. In *Journal of Phycology* 32 (5), pp. 850–857.

Chevreux, B.; Pfisterer, T.; Drescher, B.; Driesel, A. J.; Muller, W. E.; Wetter, T.; Suhai, S. (2004): Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. In *Genome research* 14 (6), pp. 1147–1159.

Clark, I. A.; Hunt, N. H. (1983): Evidence for reactive oxygen intermediates causing hemolysis and parasite death in malaria. In *Infection and Immunity* 39 (1), pp. 1–6.

Claros, M. G.; Perea, J.; Shu, Y.; Samatey, F. A.; Popot, J.-L.; Jacq, C. (1995): Limitations to in vivo Import of Hydrophobic Proteins into Yeast Mitochondria. In *The FEBS Journal* 228 (3), pp. 762–771.

Coats, D. W. (1999): Parasitic Life Styles of Marine Dinoflagellates. In *J Eukaryotic Microbiology* 46 (4), pp. 402–409.

Collen, J.; Porcel, B.; Carre, W.; Ball, S. G.; Chaparro, C.; Tonon, T. et al. (2013): Genome structure and metabolic features in the red seaweed Chondrus crispus shed light on evolution of the Archaeplastida. In *Proceedings of the National Academy of Sciences of the United States of America* 110 (13), pp. 5247–5252.

Conesa, A.; Gotz, S.; Garcia-Gomez, J. M.; Terol, J.; Talon, M.; Robles, M. (2005): Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. In *Bioinformatics (Oxford, England)* 21 (18), pp. 3674–3676.

Cordoba, E.; Salmi, M.; Leon, P. (2009): Unravelling the regulatory mechanisms that modulate the MEP pathway in higher plants. In *Journal of experimental botany* 60 (10), pp. 2933–2943.

Coss, C. A.; Robledo, J. A.; Vasta, G. R. (2001): Fine structure of clonally propagated in vitro life stages of a Perkinsus sp. isolated from the Baltic clam Macoma balthica. In *The Journal of eukaryotic microbiology* 48 (1), pp. 38–51.

Cox, F. E. G. (2002): History of Human Parasitology. In *Clinical Microbiology Reviews* 15 (4), pp. 595–612.

Crofts, A. R. (2004): The cytochrome bc1 complex: function in the context of structure. In *Annual review of physiology* 66, pp. 689–733.

Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E. (2004): WebLogo: a sequence logo generator. In *Genome research* 14 (6), pp. 1188–1190.

Cross, L. L.; Ebeed, H. T.; Baker, A. (2016): Peroxisome biogenesis, protein targeting mechanisms and PEX gene functions in plants. In *Biochimica et biophysica acta* 1863 (5), pp. 850–862.

Cumbo, Vivian R.; Baird, Andrew H.; Moore, Robert B.; Negri, Andrew P.; Neilan, Brett A.; Salih, Anya et al. (2013): Chromera velia is Endosymbiotic in Larvae of the Reef Corals Acropora digitifera and A. tenuis. In *Protist* 164 (2), pp. 237–244.

Dang, Y.; Green, B. R. (2009): Substitutional editing of Heterocapsa triquetra chloroplast transcripts and a folding model for its divergent chloroplast 16S rRNA. In *Gene* 442 (1-2), pp. 73–80.

Danne, J. C.; Gornik, S. G.; MacRae, J. I.; McConville, M. J.; Waller, R. F. (2013): Alveolate mitochondrial metabolic evolution: dinoflagellates force reassessment of the role of parasitism as a driver of change in apicomplexans. In *Molecular biology and evolution* 30 (1), pp. 123–139.

Dellibovi-Ragheb, T. A.; Gisselberg, J. E.; Prigge, S. T. (2013): Parasites FeS up: iron-sulfur cluster biogenesis in eukaryotic pathogens. In *PLoS pathogens* 9 (4), e1003227.

Delsuc, F.; Brinkmann, H.; Philippe, H. (2005): Phylogenomics and the reconstruction of the tree of life. In *Nature reviews. Genetics* 6 (5), pp. 361–375.

Delwiche, C. F. (1999): Tracing the thread of plastid diversity through the tapestry of life. In *the american naturalist* 154 (S4), S164-S177.

Des Marais, D. J. (2000): When did photosynthesis emerge on Earth? In *Science* 289 (5485), pp. 1703–1705.

Diekmann, Y.; Pereira-Leal, J. B. (2013): Evolution of intracellular compartmentalization. In *The Biochemical journal* 449 (2), pp. 319–331.

Ding, M.; Clayton, C.; Soldati, D. (2000): Toxoplasma gondii catalase: are there peroxisomes in toxoplasma? In *J Cell Sci* 113 (13), pp. 2409–2419.

Dittrich-Domergue, F.; Joubes, J.; Moreau, P.; Lessire, R.; Stymne, S.; Domergue, F. (2014): The bifunctional protein TtFARAT from Tetrahymena thermophila catalyzes the formation of both precursors required to initiate ether lipid biosynthesis. In *The Journal of biological chemistry* 289 (32), pp. 21984–21994.

Doolittle, W. F. (1998): You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. In *Trends in Genetics* 14 (8), pp. 307–311.

Dorrell, R. G.; Smith, A. G. (2011): Do red and green make brown?: perspectives on plastid acquisitions within chromalveolates. In *Eukaryotic cell* 10 (7), pp. 856–868.

Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.-T. et al. (2001): The highly reduced genome of an enslaved algal nucleus. In *Nature* 410 (6832), pp. 1091–1096.

Douglas, S. E.; Penny, S. L. (1999): The plastid genome of the cryptophyte alga, Guillardia theta: complete sequence and conserved synteny groups confirm its common ancestry with red algae. In *Journal of molecular evolution* 48 (2), pp. 236–244.

Douglas, S. E. (1998): Plastid evolution: origins, diversity, trends. In *Current opinion in genetics & development* 8 (6), pp. 655–661.

Du Yoo, Y.; Seong, K. A.; Myung, G.; Kim, H. S.; Jeong, H. J.; Palenik, B.; Yih, W. (2015): Ingestion of the unicellular cyanobacterium Synechococcus by the mixotrophic red tide ciliate Mesodinium rubrum. In *ALGAE* 30 (4), pp. 281–290.

Duhita, N.; Le, H. A. T.; Satoshi, S.; Kazuo, H.; Daisuke, M.; Takao, S. (2010): The origin of peroxisomes: The possibility of an actinobacterial symbiosis. In *Gene* 450 (1-2), pp. 18–24.

Dungan, C. F.; Bushek, D. (2015): Development and applications of Ray's fluid thioglycollate media for detection and manipulation of Perkinsus spp. pathogens of marine molluscs. In *Journal of Invertebrate Pathology* 131, pp. 68–82.

Dunthorn, M; Lipps, J. H.; Dolan, J. R.; Saab, M. A.-A.; Aescht, E.; Bachy, C. et al. (2015): Ciliates — Protists with complex morphologies and ambiguous early fossil record. In *Marine Micropaleontology* 119, pp. 1–6.

Dunthorn, M.; Lipps, J. H.; Stoeck, T. (2010): Reassessment of the putative ciliate fossils Eotintinnopsis, Wujiangella, and Yonyangella from the Neoproterozoic Doushantuo Formation in China. In *Acta Protozoologica* 2010 (2), pp. 139–144.

Duve, C. de; Baudhuin, P. (1966): Peroxisomes (microbodies and related particles). In *Physiological Reviews* 46 (2), pp. 323–357.

Dyall, S. D.; Johnson, P. J. (2000): Origins of hydrogenosomes and mitochondria: evolution and organelle biogenesis. In *Current opinion in microbiology* 3 (4), pp. 404–411.

Efremov, R. G.; Baradaran, R.; Sazanov, L. A. (2010): The architecture of respiratory complex I. In *Nature* 465 (7297), pp. 441–445.

Eisen, J. A.; Coyne, R. S.; Wu, M.; Wu, D.; Thiagarajan, M.; Wortman, J. R. et al. (2006): Macronuclear genome sequence of the ciliate Tetrahymena thermophila, a model eukaryote. In *PLoS biology* 4 (9), e286.

El Albani, A.; Bengtson, S.; Canfield, D. E.; Riboulleau, A.; Rollion Bard, C.; Macchiarelli, R. et al. (2014): The 2.1 Ga old Francevillian biota: biogenicity, taphonomy and biodiversity. In *PloS one* 9 (6), e99438.

Emanuelsson, O.; Nielsen, H.; Brunak, S.; Heijne, G. von (2000): Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. In *Journal of molecular biology* 300 (4), pp. 1005–1016.

Embley, T. M.; van der Giezen, M.; Horner, D. S.; Dyal, P. L.; Bell, S.; Foster, P. G. (2003): Hydrogenosomes, mitochondria and early eukaryotic evolution. In *IUBMB life* 55 (7), pp. 387–395.

Emelyanov, V. V. (2003): Mitochondrial connection to the origin of the eukaryotic cell. In *Eur J Biochem* 270 (8), pp. 1599–1618.

Erdmann, R.; Kunau, W. H. (1992): A genetic approach to the biogenesis of peroxisomes in the yeast Saccharomyces cerevisiae. In *Cell biochemistry and function* 10 (3), pp. 167–174.

Erdmann, R.; Veenhuis, M.; Mertens, D.; Kunau, W. H. (1989): Isolation of peroxisome-deficient mutants of Saccharomyces cerevisiae. In *Proceedings of the National Academy of Sciences* 86 (14), pp. 5419–5423.

Facchinelli, F.; Colleoni, C.; Ball, S. G.; Weber, A. P. (2013): Chlamydia, cyanobiont, or host: who was on top in the menage a trois? In *Trends in Plant Science* 18 (12), pp. 673–679.

Fagarasanu, A.; Fagarasanu, M.; Rachubinski, R. A. (2007): Maintaining peroxisome populations: a story of division and inheritance. In *Annual review of cell and developmental biology* 23, pp. 321–344.

Fang, Y.; Morrell, J. C.; Jones, J. M.; Gould, S. J. (2004): PEX3 functions as a PEX19 docking factor in the import of class I peroxisomal membrane proteins. In *The Journal of Cell Biology* 164 (6), pp. 863–875.

Fast, N. M.; Xue, L.; Bingham, S.; Keeling, P. J. (2002): Re-examining Alveolate Evolution Using Multiple Protein Molecular Phylogenies. In *J Eukaryotic Microbiology* 49 (1), pp. 30–37.

Fast, N. M.; Kissinger, J. C.; Roos, D. S.; Keeling, P. J. (2001): Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. In *Molecular biology and evolution* 18 (3), pp. 418–426.

Fayer, R.; Trout, J. M.; Jenkins, M. C. (1998): Infectivity of Cryptosporidium parvum Oocysts Stored in Water at Environmental Temperatures. In *The Journal of Parasitology* 84 (6), pp. 1165–1169.

Feagin, J. E.; Harrell, M. I.; Lee, J. C.; Coe, K. J.; Sands, B. H.; Cannone, J. J. et al. (2012): The fragmented mitochondrial ribosomal RNAs of Plasmodium falciparum. In *PloS one* 7 (6), e38320.

Feagin, J. E.; Mericle, B. L.; Werner, E.; Morris, M. (1997): Identification of additional rRNA fragments encoded by the Plasmodium falciparum 6 kb element. In *Nucleic acids research* 25 (2), pp. 438–446.

Fensome, R. A.; MacRae, R. A.; Williams, G. L. (1994): Dinoflagellate evolution and diversity through time. In *Sci Rev Bedford Inst Oceanogr* 1995, pp. 45–50.

Fernandez Robledo, J. A.; Caler, E.; Matsuzaki, M.; Keeling, P. J.; Shanmugam, D.; Roos, D. S.; Vasta, G. R. (2011): The search for the missing link: a relic plastid in Perkinsus? In *International journal for parasitology* 41 (12), pp. 1217–1229.

Flegontov, P.; Michalek, J.; Janouskovec, J.; Lai, D.-H.; Jirku, M.; Hajduskova, E. et al. (2015): Divergent mitochondrial respiratory chains in phototrophic relatives of apicomplexan parasites. In *Molecular biology and evolution* 32 (5), pp. 1115–1131.

Flügge, U.-I. (2001): Plant Chloroplasts and Other Plastids. In : Encyclopedia of Life Sciences. Chichester, UK: John Wiley & Sons, Ltd.

Frischer, M. E.; Lee, R. F.; Sheppard, M. A.; Mauer, A.; Rambow, F.; Neumann, M. et al. (2006): Evidence for a free-living life stage of the blue crab parasitic dinoflagelate, Hematodinium sp. In *Harmful Algae* 5 (5), pp. 548–557.

Fry, M.; Pudney, M. (1992): Site of action of the antimalarial hydroxynaphthoquinone, 2-[trans-4-(4'-chlorophenyl) cyclohexyl]-3- hydroxy-1,4-naphthoquinone (566C80). In *Biochemical Pharmacology* 43 (7), pp. 1545–1553.

Fujiki, Y.; Okumoto, K.; Mukai, S.; Honsho, M.; Tamura, S. (2014): Peroxisome biogenesis in mammalian cells. In *Frontiers in physiology* 5, p. 307.

Funes, S.; Davidson, E.; Reyes-Prieto, A.; Magallon, S.; Herion, P.; King, M. P.; Gonzalez-Halphen, D. (2002): A green algal apicoplast ancestor. In *Science (New York, N.Y.)* 298 (5601), p. 2155.

Furumoto, T.; Yamaguchi, T.; Ohshima-Ichie, Y.; Nakamura, M.; Tsuchida-Iwata, Y.; Shimamura, M. et al. (2011): A plastidial sodium-dependent pyruvate transporter. In *Nature* 476 (7361), pp. 472–475.

Gabaldon, T.; Ginger, M. L.; Michels, P. A. (2016): Peroxisomes in parasitic protists. In *Molecular and biochemical parasitology* (1-2), pp. 35–45.

Gabaldon, T. (2014): A metabolic scenario for the evolutionary origin of peroxisomes from the endomembranous system. In *Cellular and Molecular Life Sciences* 71 (13), pp. 2373–2376.

Gabaldon, T. (2012): Mitochondrial Origins. In E. C. Bullerwell (Ed.): Organelle Genetics: Evolution of Organelle Genomes and Gene Expression. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 3–18.

Gabaldon, T.; Capella-Gutierrez, S. (2010): Lack of phylogenetic support for a supposed actinobacterial origin of peroxisomes. In *Gene* 465 (1-2), pp. 61–65.

Gabaldon, T. (2010): Peroxisome diversity and evolution. In *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 365 (1541), pp. 765–773.

Gabaldon, T.; Snel, B.; van Zimmeren, F.; Hemrika, W.; Tabak, H.; Huynen, M. A. (2006): Origin and evolution of the peroxisomal proteome. In *Biology direct* 1 (8).

Gajadhar, A. A.; Marquardt, W. C.; Hall, R.; Gunderson, J.; Ariztia-Carmona, E. V.; Sogin, M. L. (1991): Ribosomal RNA sequences of Sarcocystis muris, Theileria annulata and Crypthecodinium cohnii reveal evolutionary relationships among apicomplexans, dinoflagellates, and ciliates. In *Molecular and biochemical parasitology* 45 (1), pp. 147–154.

Gajria, B.; Bahl, A.; Brestelli, J.; Dommer, J.; Fischer, S.; Gao, X. et al. (2008): ToxoDB: an integrated Toxoplasma gondii database resource. In *Nucleic acids research* 36 (Database issue), D553-6.

Gao, F.; Warren, A.; Zhang, Q.; Gong, J.; Miao, M.; Sun, P. et al. (2016): The All-Data-Based Evolutionary Hypothesis of Ciliated Protists with a Revised Classification of the Phylum Ciliophora (Eukaryota, Alveolata). In *Scientific reports* 6, p. 24874.

Gao, F.; Katz, L. A. (2014): Phylogenomic analyses support the bifurcation of ciliates into two major clades that differ in properties of nuclear division. In *Molecular Phylogenetics and Evolution* 70, pp. 240–243.

Gardner, M. J.; Hall, N.; Fung, E.; White, O.; Berriman, M.; Hyman, R. W. et al. (2002): Genome sequence of the human malaria parasite Plasmodium falciparum. In *Nature* 419 (6906), pp. 498–511.

Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. (2005): Protein Identification and Analysis Tools on the ExPASy Server. In J. M. Walker (Ed.): The Proteomics Protocols Handbook. Totowa, NJ: Humana Press, pp. 571–607.

Gile, G. H.; Slamovits, C. H. (2014): Transcriptomic analysis reveals evidence for a cryptic plastid in the colpodellid Voromonas pontica, a close relative of chromerids and apicomplexan parasites. In *PloS one* 9 (5), e96258.

Gilson, P. R.; Su, V.; Slamovits, C. H.; Reith, M. E.; Keeling, P. J.; McFadden, G. I. (2006): Complete nucleotide sequence of the chlorarachniophyte nucleomorph: nature's smallest nucleus. In *Proceedings of the National Academy of Sciences of the United States of America* 103 (25), pp. 9566–9571.

Ginger, M. L. (2006): Niche metabolism in parasitic protozoa. In *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 361 (1465), pp. 101–118.

Gisselberg, J. E.; Dellibovi-Ragheb, T. A.; Matthews, K. A.; Bosch, G.; Prigge, S. T. (2013): The suf iron-sulfur cluster synthesis pathway is required for apicoplast maintenance in malaria parasites. In *PLoS pathogens* 9 (9), e1003655.

Glazer, A. (1985): Light Harvesting by Phycobilisomes. In *Annual Review of Biophysics and Biomolecular Structure* 14 (1), pp. 47–77.

Gomez, F.; Moreira, D.; Lopez-Garcia, P. (2010): Molecular phylogeny of noctilucoid dinoflagellates (Noctilucales, Dinophyceae). In *Protist* 161 (3), pp. 466–478.

Gonzalez, N. H.; Felsner, G.; Schramm, F. D.; Klingl, A.; Maier, U.-G.; Bolte, K. (2011): A single peroxisomal targeting signal mediates matrix protein import in diatoms. In *PloS one* 6 (9), e25316.

Gornik, S. G.; Febrimarsa; Cassin, A. M.; MacRae, J. I.; Ramaprasad, A.; Rchiad, Z. et al. (2015): Endosymbiosis undone by stepwise elimination of the plastid in a parasitic dinoflagellate. In *Proceedings of the National Academy of Sciences of the United States of America* 112 (18), pp. 5767–5772.

Graaf, R. M. de; van Alen, T. A.; Dutilh, B. E.; Kuiper, J. W. P.; van Zoggel, H. J.; Huynh, M. B. et al. (2009): The mitochondrial genomes of the ciliates Euplotes minuta and Euplotes crassus. In *BMC genomics* 10, p. 514.

Graham, L. E.; Wilcox, L. W. (2000): Algae. Upper Saddler River, NJ: Prentice Hall.

Graindorge, A.; Frenal, K.; Jacot, D.; Salamun, J.; Marq, J. B.; Soldati-Favre, D. (2016): The Conoid Associated Motor MyoH Is Indispensable for Toxoplasma gondii Entry and Exit from Host Cells. In *PLoS pathogens* 12 (1), e1005388.

Grauvogel, C.; Reece, K. S.; Brinkmann, H.; Petersen, J. (2007): Plastid isoprenoid metabolism in the oyster parasite Perkinsus marinus connects dinoflagellates and malaria pathogens--new impetus for studying alveolates. In *Journal of molecular evolution* 65 (6), pp. 725–729.

Gray, M. W. (2015): Mosaic nature of the mitochondrial proteome: Implications for the origin and evolution of mitochondria. In *Proceedings of the National Academy of Sciences of the United States of America* 112 (33), pp. 10133–10138.

Gray, M. W.; Lang, B. F.; Burger, G. (2004): Mitochondria of protists. In *Annual review of genetics* 38, pp. 477–524.

Gray, M. W. (2003): Diversity and evolution of mitochondrial RNA editing systems. In *IUBMB life* 55 (4-5), pp. 227–233.

Gray, M. W.; Burger, G.; Lang, B. F. (2001): The origin and early evolution of mitochondria. In *Genome biology* 2 (6), reviews1018.1–1018.5.

Gray, M. W. (1999): Evolution of organellar genomes. In *Current opinion in genetics & development* 9 (6), pp. 678–687.

Hackett, J. D.; Yoon, H. S.; Li, S.; Reyes-Prieto, A.; Rummele, S. E.; Bhattacharya, D. (2007): Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of rhizaria with chromalveolates. In *Molecular biology and evolution* 24 (8), pp. 1702–1713.

Hackett, J. D.; Yoon, H. S.; Soares, M. B.; Bonaldo, M. F.; Casavant, T. L.; Scheetz, T. E. et al. (2004): Migration of the plastid genome to the nucleus in a peridinin dinoflagellate. In *Current biology : CB* 14 (3), pp. 213–218.

Hackett, J. D.; Anderson, D. M.; Erdner, D. L.; Bhattacharya, D. (2004): Dinoflagellates: a remarkable evolutionary experiment. In *American journal of botany* 91 (10), pp. 1523–1534.

Hackett, J. D.; Maranda, L.; Yoon, H. S.; Bhattacharya, D. (2003): Phylogenetic evidence for the cryptophyte origin of the plastid of Dinophysis (Dinophysiales, Dinophyceae) 1. In *Journal of Phycology* 39 (2), pp. 440–448.

Hackstein, J. H. P.; Tjaden, J.; Huynen, M. (2006): Mitochondria, hydrogenosomes and mitosomes: products of evolutionary tinkering! In *Current genetics* 50 (4), pp. 225–245.

Hall, T. A. (1999): BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In *Nucleic Acids Symposium Series* 41, pp. 95–98.

Hampl, V.; Hug, L.; Leigh, J. W.; Dacks, J. B.; Lang, B. F.; Simpson, A. G. B.; Roger, A. J. (2009): Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". In *Proceedings of the National Academy of Sciences of the United States of America* 106 (10), pp. 3859–3864.

Hansen, P. J.; Moldrup, M.; Tarangkoon, W.; Garcia-Cuetos, L.; Moestrup, Ø. (2012): Direct evidence for symbiont sequestration in the marine red tide ciliate Mesodinium rubrum. In *Aquat. Microb. Ecol.* 66 (1), pp. 63–75.

Hansen, F. C.; Witte, H. J.; Passarge, J. (1996): Grazing in the heterotrophic dinoflagellate Oxyrrhis marina: size selectivity and preference for calcified Emiliania huxleyi cells. In *Aquatic Microbial Ecology* 10 (3), pp. 307–313.

Harper, J. T.; Waanders, E.; Keeling, P. J. (2005): On the monophyly of chromalveolates using a six-protein phylogeny of eukaryotes. In *International journal of systematic and evolutionary microbiology* 55 (Pt 1), pp. 487–496.

Harper, J. T.; Keeling, P. J. (2003): Nucleus-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids. In *Molecular biology and evolution* 20 (10), pp. 1730–1735.

Hausmann, K. (2002): Food acquisition, food ingestion and food digestion by protists. In *Jpn. J. Protozool. Vol* 35 (2), pp. 85–95.

Hayashi, M.; Toriyama, K.; Kondo, M.; Kato, A.; Mano, S.; Bellis, L. de et al. (2000): Functional Transformation of Plant Peroxisomes. In *CBB* 32 (1-3), pp. 295–304.

Heiges, M.; Wang, H.; Robinson, E.; Aurrecoechea, C.; Gao, X.; Kaluskar, N. et al. (2006): CryptoDB: a Cryptosporidium bioinformatics resource update. In *Nucleic acids research* 34 (Database issue), D419-22.

Heijne, G. von (1986): Why mitochondria need a genome. In *FEBS letters* 198 (1), pp. 1–4.

Hempel, F.; Bullmann, L.; Lau, J.; Zauner, S.; Maier, U. G. (2009): ERAD-derived preprotein transport across the second outermost plastid membrane of diatoms. In *Molecular biology and evolution* 26 (8), pp. 1781–1790.

Henriquez, F. L.; Richards, T. A.; Roberts, F.; McLeod, R.; Roberts, C. W. (2005): The unusual mitochondrial compartment of Cryptosporidium parvum. In *Trends in parasitology* 21 (2), pp. 68–74.

Hikosaka, K.; Nakai, Y.; Watanabe, Y.-I.; Tachibana, S.-I.; Arisue, N.; Palacpac, N. M. Q. et al. (2011): Concatenated mitochondrial DNA of the coccidian parasite Eimeria tenella. In *Mitochondrion* 11 (2), pp. 273–278.

Hikosaka, K.; Watanabe, Y.-I.; Tsuji, N.; Kita, K.; Kishine, H.; Arisue, N. et al. (2010): Divergence of the mitochondrial genome structure in the apicomplexan parasites, Babesia and Theileria. In *Molecular biology and evolution* 27 (5), pp. 1107–1116.

Hiller, R. G. (2001): 'Empty' minicircles and petB/atpA and psbD/psbE (cyt b559 α) genes in tandem in Amphidinium carterae plastid DNA1. In *FEBS letters* 505 (3), pp. 449–452.

Hirokawa, T.; Boon-Chieng, S.; Mitaku, S. (1998): SOSUI: classification and secondary structure prediction system for membrane proteins. In *Bioinformatics (Oxford, England)* 14 (4), pp. 378–379.

Hoek, C.; Mann, D.; Jahns, H. M. (1995): Algae: an introduction to phycology: Cambridge university press.

Hoepfner, D.; Schildknegt, D.; Braakman, I.; Philippsen, P.; Tabak, H. F. (2005): Contribution of the endoplasmic reticulum to peroxisome formation. In *Cell* 122 (1), pp. 85–95.

Hogg, J. F. (1969): Peroxisomes in Tetrahymena and their relation to gluconeogenesis. In *Annals of the New York Academy of Sciences* 168 (2), pp. 281–291.

Hogg, J. F.; Kornberg, H. L. (1963): The metabolism of C2-compounds in micro-organisms. 9. Role of the glyoxylate cycle in protozoal glyconeogenesis. In *Biochemical Journal* 86 (3), p. 462.

Hohmann-Marriott, M. F.; Blankenship, R. E. (2011): Evolution of photosynthesis. In *Annual review of plant biology* 62, pp. 515–548.

Holland, H. D. (2002): Volcanic gases, black smokers, and the great oxidation event. In *Geochimica et Cosmochimica Acta* 66 (21), pp. 3811–3826.

Honsho, M.; Hiroshige, T.; Fujiki, Y. (2002): The membrane biogenesis peroxin Pex16p. Topogenesis and functional roles in peroxisomal membrane assembly. In *The Journal of biological chemistry* 277 (46), pp. 44513–44524.

Hoppenrath, M.; Leander, B. S. (2010): Dinoflagellate phylogeny as inferred from heat shock protein 90 and ribosomal gene sequences. In *PloS one* 5 (10), e13220.

Hroudova, J.; Singh, N.; Fisar, Z. (2014): Mitochondrial dysfunctions in neurodegenerative diseases: relevance to Alzheimer's disease. In *BioMed research international* 2014, p. 175062.

Hsu, E. (2006): Reflections on the 'discovery' of the antimalarial qinghao. In *British journal of clinical pharmacology* 61 (6), pp. 666–670.

Huang, J.; Gogarten, J. P. (2007): Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? In *Genome biology* 8 (6), R99.

Huang, J.; Mullapudi, N.; Lancto, C. A.; Scott, M.; Abrahamsen, M. S.; Kissinger, J. C. (2004): Phylogenomic evidence supports past endosymbiosis, intracellular and horizontal gene transfer in Cryptosporidium parvum. In *Genome biology* 5 (11), R88.

Hunter, W. N. (2011): Isoprenoid precursor biosynthesis offers potential targets for drug discovery against diseases caused by apicomplexan parasites. In *Current topics in medicinal chemistry* 11 (16), pp. 2048–2059.

Hunter, W. N. (2007): The non-mevalonate pathway of isoprenoid precursor biosynthesis. In *The Journal of biological chemistry* 282 (30), pp. 21573–21577.

Imanian, B.; Pombert, J.-F.; Dorrell, R. G.; Burki, F.; Keeling, P. J. (2012): Tertiary endosymbiosis in two dinotoms has generated little change in the mitochondrial genomes of their dinoflagellate hosts and diatom endosymbionts. In *PloS one* 7 (8), e43763.

Inagaki, Y.; Dacks, J. B.; Doolittle, W. Ford; Watanabe, K. I.; Ohama, T. (2000): Evolutionary relationship between dinoflagellates bearing obligate diatom endosymbionts: insight into tertiary endosymbiosis. In *International journal of systematic and evolutionary microbiology* 50 (6), pp. 2075–2081.

Islinger, M.; Cardoso, M. J. R.; Schrader, M. (2010): Be different--the diversity of peroxisomes in the animal kingdom. In *Biochimica et biophysica acta* 1803 (8), pp. 881–897.

Iverson, T. M. (2013): Catalytic mechanisms of complex II enzymes: a structural perspective. In *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1827 (5), pp. 648–657.

Jackson, C. J.; Waller, R. F. (2013): A widespread and unusual RNA trans-splicing type in dinoflagellate mitochondria. In *PloS one* 8 (2), e56777.

Jackson, C. J.; Gornik, S. G.; Waller, R. F. (2012): The mitochondrial genome and transcriptome of the basal dinoflagellate Hematodinium sp.: character evolution within the highly derived mitochondrial genomes of dinoflagellates. In *Genome biology and evolution* 4 (1), pp. 59–72.

Jackson, C. J.; Norman, J. E.; Schnare, M. N.; Gray, M. W.; Keeling, P. J.; Waller, R. F. (2007): Broad genomic and transcriptional analysis reveals a highly derived genome in dinoflagellate mitochondria. In *BMC biology* 5, p. 41.

Janouskovec, J.; Tikhonenkov, D. V.; Burki, F.; Howe, A. T.; Kolisko, M.; Mylnikov, A. P.; Keeling, P. J. (2015): Factors mediating plastid dependency and the origins of parasitism in apicomplexans and their close relatives. In *Proceedings of the National Academy of Sciences of the United States of America* 112 (33), pp. 10200–10207.

Janouskovec, J.; Liu, S.-L.; Martone, P. T.; Carre, W.; Leblanc, C.; Collen, J.; Keeling, P. J. (2013): Evolution of red algal plastid genomes: ancient architectures, introns, horizontal gene transfer, and taxonomic utility of plastid markers. In *PloS one* 8 (3), e59001.

Janouskovec, J.; Horak, A.; Obornik, M.; Lukes, J.; Keeling, P. J. (2010): A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. In *Proceedings of the National Academy of Sciences of the United States of America* 107 (24), pp. 10949–10954.

Johnson, M. D.; Stoecker, D. K. (2005): Role of feeding in growth and photophysiology of Myrionecta rubra. In *Aquat. Microb. Ecol.* 39, pp. 303–312.

Jomaa, H.; Wiesner, J.; Sanderbrand, S.; Altincicek, B.; Weidemeyer, C.; Hintz, M. et al. (1999): Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. In *Science (New York, N.Y.)* 285 (5433), pp. 1573–1576.

Jones, J. M.; Morrell, J. C.; Gould, S. J. (2004): PEX19 is a predominantly cytosolic chaperone and import receptor for class 1 peroxisomal membrane proteins. In *The Journal of Cell Biology* 164 (1), pp. 57–67

Joseph, S. J.; Fernandez-Robledo, J. A.; Gardner, M. J.; El-Sayed, N. M.; Kuo, C.-H.; Schott, E. J. et al. (2010): The Alveolate Perkinsus marinus: biological insights from EST gene discovery. In *BMC genomics* 11, p. 228.

Kaasch, A. J.; Joiner, K. A. (2000): Targeting and Subcellular Localization of Toxoplasma gondii Catalase. Identification of Peroxisomes in an Apicomplexan Parasite. In *Journal of Biological Chemistry* 275 (2), pp. 1112–1118.

Kairo, A.; Fairlamb, A. H.; Gobright, E.; Nene, V. (1994): A 7.1 kb linear DNA molecule of Theileria parva has scrambled rDNA sequences and open reading frames for mitochondrially encoded proteins. In *The EMBO journal* 13 (4), p. 898.

Kanehisa, M.; Goto, S. (2000): KEGG: kyoto encyclopedia of genes and genomes. In *Nucleic acids research* 28 (1), pp. 27–30.

Kaneshiro, E. S.; Sul, D.; Hazra, B. (2000): Effects of Atovaquone and Diospyrin-Based Drugs on Ubiquinone Biosynthesis in Pneumocystis carinii Organisms. In *Antimicrobial agents and chemotherapy* 44 (1), pp. 14–18.

Katris, N. J.; van Dooren, G. G.; McMillan, P. J.; Hanssen, E.; Tilley, L.; Waller, R. F. (2014): The apical complex provides a regulated gateway for secretion of invasion factors in Toxoplasma. In *PLoS pathogens* 10 (4), e1004074.

Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S. et al. (2012): Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. In *Bioinformatics (Oxford, England)* 28 (12), pp. 1647–1649.

Keeling, P. J. (2010): The endosymbiotic origin, diversification and fate of plastids. In *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 365 (1541), pp. 729–748.

Keeling, P. J. (2004): Diversity and evolutionary history of plastids and their hosts. In *American journal of botany* 91 (10), pp. 1481–1493.

Kerscher, S. J. (2000): Diversity and origin of alternative NADH:ubiquinone oxidoreductases. In *Biochimica et biophysica acta* 1459 (2-3), pp. 274–283.

Kim, E.; Archibald, J. M. (2009): Diversity and Evolution of Plastids and Their Genomes: Springer Berlin Heidelberg.

Kim, S. H.; Vlkolinsky, R.; Cairns, N.; Lubec, G. (2000): Decreased levels of complex III core protein 1 and complex V β chain in brains from patients with Alzheimer's disease and Down syndrome. In *CMLS, Cell. Mol. Life Sci.* 57 (12), pp. 1810–1816.

Kohli, G. S.; John, U.; van Dolah, F. M.; Murray, S. A. (2016): Evolutionary distinctiveness of fatty acid and polyketide synthesis in eukaryotes. In *The ISME journal*, pp. 1877–1890.

Kopylova, E.; Noe, L.; Touzet, H. (2012): SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. In *Bioinformatics (Oxford, England)* 28 (24), pp. 3211–3217.

Koreny, L.; Sobotka, R.; Janouskovec, J.; Keeling, P. J.; Obornik, M. (2011): Tetrapyrrole synthesis of photosynthetic chromerids is likely homologous to the unusual pathway of apicomplexan parasites. In *The Plant cell* 23 (9), pp. 3454–3462.

Kornberg, H. L.; Krebs, H. A. (1957): Synthesis of cell constituents from C2-units by a modified tricarboxylic acid cycle. In *Nature* 179 (4568), p. 988.

Koumandou, V. L.; Nisbet, R. E. R.; Barbrook, A. C.; Howe, C. J. (2004): Dinoflagellate chloroplasts-where have all the genes gone? In *Trends in genetics : TIG* 20 (5), pp. 261–267.

Kunze, M.; Hartig, A. (2013): Permeability of the peroxisomal membrane: lessons from the glyoxylate cycle. In *Frontiers in physiology* 4, p. 204.

Kunze, M.; Pracharoenwattana, I.; Smith, S. M.; Hartig, A. (2006): A central role for the peroxisomal membrane in glyoxylate cycle function. In *Biochimica et biophysica acta* 1763 (12), pp. 1441–1452.

Kuo, C.-H.; Wares, J. P.; Kissinger, J. C. (2008): The Apicomplexan whole-genome phylogeny: an analysis of incongruence among gene trees. In *Molecular biology and evolution* 25 (12), pp. 2689–2698.

Kurland, C. G.; Andersson, S. G. E. (2000): Origin and Evolution of the Mitochondrial Proteome. In *Microbiology and Molecular Biology Reviews* 64 (4), pp. 786–820.

Laatsch, T.; Zauner, S.; Stoebe-Maier, B.; Kowallik, K. V.; Maier, U-G (2004): Plastid-derived single gene minicircles of the dinoflagellate Ceratium horridum are localized in the nucleus. In *Molecular biology and evolution* 21 (7), pp. 1318–1322.

LaGier, M. J.; Tachezy, J.; Stejskal, F.; Kutisova, K.; Keithly, J. S. (2003): Mitochondrial-type iron-sulfur cluster biosynthesis genes (IscS and IscU) in the apicomplexan Cryptosporidium parvum. In *Microbiology (Reading, England)* 149 (Pt 12), pp. 3519–3530.

LaJeunesse, T. C.; Lambert, G.; Andersen, R. A.; Coffroth, M. A.; Galbraith, D. W. (2005): Symbiodinium (Pyrrophyta) genome sizes (DNA content) are smallest among dinoflagellates. In *J Phycol* 41 (4), pp. 880–886.

Lambert, D. H.; Bryant, D. A.; Stirewalt, V. L.; Dubbs, J. M.; Stevens, S. E.; Porter, R. D. (1985): Gene map for the Cyanophora paradoxa cyanelle genome. In *Journal of bacteriology* 164 (2), pp. 659–664.

Lametschwandtner, G. (1998): The Difference in Recognition of Terminal Tripeptides as Peroxisomal Targeting Signal 1 between Yeast and Human Is Due to Different Affinities of Their Receptor Pex5p to the Cognate Signal and to Residues Adjacent to It. In *Journal of Biological Chemistry* 273 (50), pp. 33635–33643.

Lange, B. M.; Rujan, T.; Martin, W.; Croteau, R. (2000): Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. In *Proceedings of the National Academy of Sciences of the United States of America* 97 (24), pp. 13172–13177.

Lartillot, N.; Philippe, H. (2004): A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. In *Molecular biology and evolution* 21 (6), pp. 1095–1109.

Lau, A. O. T. (2009): An overview of the Babesia, Plasmodium and Theileria genomes: a comparative perspective. In *Molecular and biochemical parasitology* 164 (1), pp. 1–8.

Lazarow, P. B.; Fujiki, Y. (1985): Biogenesis of peroxisomes. In *Annual review of cell biology* 1 (1), pp. 489–530.

Le, S. Q.; Gascuel, O. (2008): An improved general amino acid replacement matrix. In *Molecular biology and evolution* 25 (7), pp. 1307–1320.

Le, S. Q.; Lartillot, N.; Gascuel, O. (2008): Phylogenetic mixture models for proteins. In *Philosophical Transactions of the Royal Society B: Biological Sciences* 363 (1512), pp. 3965–3976.

Le Herisse, A.; Masure, E.; Javaux, E. J.; Marshall, C. P. (2012): The end of a myth. Arpylorus antiquus paleozoic dinoflagellate cyst. In *PALAIOS* 27 (6), pp. 414–423.

Leander, B. S.; Keeling, P. J. (2004): Early Evolutionary History of Dinoflagellates and Apicomplexans (Alveolata) as Inferred from HSP90 and Actin Phylogenies. In *Journal of Phycology* 40 (2), pp. 341–350.

Leander, B. S.; Clopton, R. E.; Keeling, P. J. (2003): Phylogeny of gregarines (Apicomplexa) as inferred from small-subunit rDNA and beta-tubulin. In *International journal of systematic and evolutionary microbiology* 53 (Pt 1), pp. 345–354.

Leander, B. S.; Triemer, R. E.; Farmer, M. A. (2001): Character evolution in heterotrophic euglenids. In *European Journal of Protistology* 37 (3), pp. 337–356.

Levine, N. D. (1971): Uniform Terminology for the Protozoan Subphylum Apicomplexa. In *The Journal of Protozoology* 18 (2), pp. 352–355.

Li, C-W; Chen, J-Y; Lipps, J. H.; Gao, F.; Chi, H-M; Wu, H-J (2007): Ciliated protozoans from the precambrian doushantuo formation, Wengan, South China. In *Geological Society, London, Special Publications* 286 (1), pp. 151–156.

Li, W.; Mo, W.; Shen, D.; Sun, L.; Wang, J.; Lu, S. et al. (2005): Yeast model uncovers dual roles of mitochondria in action of artemisinin. In *PLoS genetics* 1 (3), e36.

Li, Y.-X.; Zhang, S.-X.; Zhang, J. (2009): Mesoproterozoic calymmian tintinnids from central China. In *The Open Paleontology Journal* 2 (1), pp. 10–13.

Lichtenthaler, H. K. (2000): Non-mevalonate isoprenoid biosynthesis: enzymes, genes and inhibitors. In *Biochemical Society transactions* 28 (6), pp. 785–789.

Lill, R. (2009): Function and biogenesis of iron-sulphur proteins. In *Nature* 460 (7257), pp. 831–838.

Lill, R.; Mühlenhoff, U. (2006): Iron-sulfur protein biogenesis in eukaryotes: components and mechanisms. In *Annual review of cell and developmental biology* 22, pp. 457–486.

Lill, R.; Kispal, G. (2000): Maturation of cellular Fe–S proteins: an essential function of mitochondria. In *Trends in biochemical sciences* 25 (8), pp. 352–356.

Lim, L.; McFadden, G. I. (2010): The evolution, metabolism and functions of the apicoplast. In *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 365 (1541), pp. 749–763.

Lin, S. (2011): Genomic understanding of dinoflagellates. In *Research in microbiology* 162 (6), pp. 551–569.

Lin, S. S.; Gross, U.; Bohne, W. (2011): Two internal type II NADH dehydrogenases of Toxoplasma gondii are both required for optimal tachyzoite growth. In *Molecular microbiology* 82 (1), pp. 209–221.

Lindmark, D. G.; Müller, M. (1973): Hydrogenosome, a cytoplasmic organelle of the anaerobic flagellate Tritrichomonas foetus, and its role in pyruvate metabolism. In *Journal of Biological Chemistry* 248 (22), pp. 7724–7728.

Lismont, C.; Nordgren, M.; van Veldhoven, P. P.; Fransen, M. (2015): Redox interplay between mitochondria and peroxisomes. In *Frontiers in Cell and Developmental Biology* 3, p. 35.

Loeblich, A. R.; Loeblich, L. A. (1985): Dinoflagellates: structure of the amphiesma and re-analysis of thecal plate patterns. In *Hydrobiologia* 123 (2), pp. 177–179.

Lohr, M.; Schwender, J.; Polle, J. E. W. (2012): Isoprenoid biosynthesis in eukaryotic phototrophs: a spotlight on algae. In *Plant science : an international journal of experimental plant biology* 185-186, pp. 9–22.

Lohse, M.; Drechsel, O.; Kahlau, S.; Bock, R. (2013): OrganellarGenomeDRAW--a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. In *Nucleic acids research* 41 (Web Server issue), W575-81.

Mai, Z.; Ghosh, S.; Frisardi, M.; Rosenthal, B.; Rogers, R.; Samuelson, J. (1999): Hsp60 Is Targeted to a Cryptic Mitochondrion-Derived Organelle ("Crypton") in the Microaerophilic Protozoan Parasite Entamoeba histolytica. In *Mol. Cell. Biol.* 19 (3), pp. 2198–2205.

Marande, W.; Burger, G. (2007): Mitochondrial DNA as a genomic jigsaw puzzle. In *Science (New York, N.Y.)* 318 (5849), p. 415.

Marande, W.; Lukeš, J.; Burger, G. (2005): Unique Mitochondrial Genome Structure in Diplonemids, the Sister Group of Kinetoplastids. In *Eukaryotic Cell* 4 (6), pp. 1137–1146.

Margulis, L. (1971): The Origin of Plant and Animal Cells: The serial symbiosis view of the origin of higher cells suggests that the customary division of living things into two kingdoms should be reconsidered. In *American scientist* 59 (2), pp. 230–235.

Markham, P.; Collinge, A. J. (1987): Woronin bodies of filamentous fungi. In *FEMS Microbiology Letters* 46 (1), pp. 1–11.

Martin, W. (2003): Gene transfer from organelles to the nucleus: frequent and in big chunks. In *Proceedings of the National Academy of Sciences of the United States of America* 100 (15), pp. 8612–8614.

Martin, W.; Herrmann, R. G. (1998): Gene transfer from organelles to the nucleus: how much, what happens, and why? In *Plant physiology* 118 (1), pp. 9–17.

Martin, W.; Müller, M. (1998): The hydrogen hypothesis for the first eukaryote. In *Nature* 392 (6671), pp. 37–41.

Martin, W.; Schnarrenberger, C. (1997): The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis. In *Current genetics* 32 (1), pp. 1–18.

Martin-Gonzalez, A.; Wierzchos, J.; Gutierrez, J. C.; Alonso, J.; Ascaso, C. (2008): Morphological stasis of protists in lower cretaceous amber. In *Protist* 159 (2), pp. 251–257.

Masuda, I.; Matsuzaki, M.; Kita, K. (2010): Extensive frameshift at all AGG and CCC codons in the mitochondrial cytochrome c oxidase subunit 1 gene of Perkinsus marinus (Alveolata; Dinoflagellata). In *Nucleic acids research* 38 (18), pp. 6186–6194.

Matsuzaki, M.; Kuroiwa, H.; Kuroiwa, T.; Kita, K.; Nozaki, H. (2008): A cryptic algal group unveiled: a plastid biosynthesis pathway in the oyster parasite Perkinsus marinus. In *Molecular biology and evolution* 25 (6), pp. 1167–1179.

Maxwell, D. P.; Wang, Y.; McIntosh, L. (1999): The alternative oxidase lowers mitochondrial reactive oxygen production in plant cells. In *Proceedings of the National Academy of Sciences of the United States of America* 96 (14), pp. 8271–8276.

McBride, H. M.; Neuspiel, M.; Wasiak, S. (2006): Mitochondria: more than just a powerhouse. In *Current biology : CB* 16 (14), R551-60.

McCammon, M. T.; Veenhuis, M.; Trapp, S. B.; Goodman, J. M. (1990): Association of glyoxylate and beta-oxidation enzymes with peroxisomes of Saccharomyces cerevisiae. In *Journal of bacteriology* 172 (10), pp. 5816–5827.

McFadden, G. I. (2011): The apicoplast. In *Protoplasma* 248 (4), pp. 641–650.

McIntosh, L. (1994): Molecular biology of the alternative oxidase. In *Plant physiology* 105 (3), pp. 781–786.

Meinecke, M.; Cizmowski, C.; Schliebs, W.; Kruger, V.; Beck, S.; Wagner, R.; Erdmann, R. (2010): The peroxisomal importomer constitutes a large and highly dynamic pore. In *Nature cell biology* 12 (3), pp. 273–277.

Mereschkowsky, C. (1905): Über Natur und Ursprung der Chromatophoren im Pflanzenreiche, Biologisches Centralblatt 25. In *J. Rosenthal*, pp. 38–604.

Mitchison, T. J. (2010): Remaining Mysteries of the Cytoplasm. In *Molecular Biology of the Cell* 21 (22), pp. 3811–3812.

Miyadera, H.; Shiomi, K.; Ui, H.; Yamaguchi, Y.; Masuma, R.; Tomoda, H. et al. (2003): Atpenins, potent and specific inhibitors of mitochondrial complex II (succinate-ubiquinone oxidoreductase). In *Proceedings of the National Academy of Sciences of the United States of America* 100 (2), pp. 473–477.

Mogi, T.; Kita, K. (2010): Diversity in mitochondrial metabolic pathways in parasitic protists Plasmodium and Cryptosporidium. In *Parasitology international* 59 (3), pp. 305–312.

Molina, J.; Hazzouri, K. M.; Nickrent, D.; Geisler, M.; Meyer, R. S.; Pentony, M. M. et al. (2014): Possible loss of the chloroplast genome in the parasitic flowering plant Rafflesia lagascae (Rafflesiaceae). In *Molecular biology and evolution* 31 (4), pp. 793–803.

Moore, R. B.; Obornik, M.; Janouskovec, J.; Chrudimsky, T.; Vancova, M.; Green, D. H. et al. (2008): A photosynthetic alveolate closely related to apicomplexan parasites. In *Nature* 451 (7181), pp. 959–963.

Moradian, M. M.; Beglaryan, D.; Skozylas, J. M.; Kerikorian, V. (2007): Complete mitochondrial genome sequence of three Tetrahymena species reveals mutation hot spots and accelerated nonsynonymous substitutions in Ymf genes. In *PloS one* 2 (7), e650.

Moran, N. A. (2002): Microbial Minimalism: Genome Reduction in Bacterial Pathogens. In *Cell* 108 (5), pp. 583–586.

Morden, C. W.; Sherwood, A. R. (2002): Continued evolutionary surprises among dinoflagellates. In *Proceedings of the National Academy of Sciences of the United States of America* 99 (18), pp. 11558–11560.

Moriya, Y.; Itoh, M.; Okuda, S.; Yoshizawa, A. C.; Kanehisa, M. (2007): KAAS: an automatic genome annotation and pathway reconstruction server. In *Nucleic acids research* 35 (Web Server issue), W182-5.

Morse, D.; Salois, P.; Markovic, P.; Hastings, J. W. (1995): A nuclear-encoded form II RuBisCO in dinoflagellates. In *Science* 268 (5217), pp. 1622–1624.

Moszczynski, K.; Mackiewicz, P.; Bodyl, A. (2012): Evidence for horizontal gene transfer from bacteroidetes bacteria to dinoflagellate minicircles. In *Molecular biology and evolution* 29 (3), pp. 887–892.

Moustafa, A.; Beszteri, B.; Maier, U. G.; Bowler, C.; Valentin, K.; Bhattacharya, D. (2009): Genomic footprints of a cryptic plastid endosymbiosis in diatoms. In *Science (New York, N.Y.)* 324 (5935), pp. 1724–1726.

Mukherjee, A.; Sadhukhan, G. C. (2016): Anti-malarial Drug Design by Targeting Apicoplasts: New Perspectives. In *Journal of pharmacopuncture* 19 (1), pp. 7–15.

Müller, M.; Martin, W. (1999): The genome of Rickettsia prowazekii and some thoughts on the origin of mitochondria and hydrogenosomes. In *Bioessays* 21 (5), pp. 377–381.

Müller, M. (1993): Review Article: The hydrogenosome. In *Microbiology* 139 (12), pp. 2879–2889.

Müller, M. (1973): Peroxisomes and hydrogenosomes in protozoa. In *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* 21 (11), pp. 955–957.

Mungpakdee, S.; Shinzato, C.; Takeuchi, T.; Kawashima, T.; Koyanagi, R.; Hisata, K. et al. (2014): Massive gene transfer and extensive RNA editing of a symbiotic dinoflagellate plastid genome. In *Genome biology and evolution* 6 (6), pp. 1408–1422.

Nakai, K.; Horton, P. (1999): PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. In *Trends in biochemical sciences* 24 (1), pp. 34–36.

Nash, E. A.; Nisbet, R. E. R.; Barbrook, A. C.; Howe, C. J. (2008): Dinoflagellates: a mitochondrial genome all at sea. In *Trends in genetics : TIG* 24 (7), pp. 328–335.

Nash, E. A.; Barbrook, A. C.; Edwards-Stuart, R. K.; Bernhardt, K.; Howe, C. J.; Nisbet, R. E. R. (2007): Organization of the mitochondrial genome in the dinoflagellate Amphidinium carterae. In *Molecular biology and evolution* 24 (7), pp. 1528–1536.

Nielsen, H.; Krogh, A. (1998): Prediction of signal peptides and signal anchors by a hidden Markov model. In *Proceedings / ... International Conference on Intelligent Systems for Molecular Biology ; ISMB. International Conference on Intelligent Systems for Molecular Biology* 6, pp. 122–130.

Nielsen, H.; Engelbrecht, J.; Brunak, S.; Heijne, G. von (1997): Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. In *Protein engineering* 10 (1), pp. 1–6.

Nisbet, R. E. R.; Koumandou, L. V.; Barbrook, A. C.; Howe, C. J. (2004): Novel plastid gene minicircles in the dinoflagellate Amphidinium operculatum. In *Gene* 331, pp. 141–147.

Noffke, N.; Christian, D.; Wacey, D.; Hazen, R. M. (2013): Microbially induced sedimentary structures recording an ancient ecosystem in the ca. 3.48 billion-year-old Dresser Formation, Pilbara, Western Australia. In *Astrobiology* 13 (12), pp. 1103–1124.

Norman, J. E.; Gray, M. W. (2001): A complex organization of the gene encoding cytochrome oxidase subunit 1 in the mitochondrial genome of the dinoflagellate, Crypthecodinium cohnii: homologous recombination generates two different cox1 open reading frames. In *Journal of molecular evolution* 53 (4-5), pp. 351–363.

Nosenko, T.; Lidie, K. L.; van Dolah, F. M.; Lindquist, E.; Cheng, J. F.; Bhattacharya, D. (2006): Chimeric plastid proteome in the Florida "red tide" dinoflagellate Karenia brevis. In *Molecular biology and evolution* 23 (11), pp. 2026–2038.

Obornik, M.; Modry, D.; Lukes, M.; Cernotikova-Stribrna, E.; Cihlar, J.; Tesarova, M. et al. (2012): Morphology, ultrastructure and life cycle of Vitrella brassicaformis n. sp., n. gen., a novel chromerid from the Great Barrier Reef. In *Protist* 163 (2), pp. 306–323.

Okamoto, N.; Horak, A.; Keeling, P. J. (2012): Description of two species of early branching dinoflagellates, Psammosa pacifica n. g., n. sp. and P. atlantica n. sp. In *PloS one* 7 (6), e34900.

Opperdoes, F. R.; Borst, P. (1977): Localization of nine glycolytic enzymes in a microbody-like organelle in Trypanosoma brucei: The glycosome. In *FEBS letters* 80 (2), pp. 360–364.

Orr, R. J. S.; Murray, S. A.; Stuken, A.; Rhodes, L.; Jakobsen, K. S. (2012): When naked became armored: an eight-gene phylogeny reveals monophyletic origin of theca in dinoflagellates. In *PloS one* 7 (11), e50004.

Ossorio, P. N.; Sibley, L. D.; Boothroyd, J. C. (1991): Mitochondrial-like DNA sequences flanked by direct and inverted repeats in the nuclear genome of Toxoplasma gondii. In *Journal of molecular biology* 222 (3), pp. 525–536.

Outten, F. W. (2015): Recent advances in the Suf Fe-S cluster biogenesis pathway: Beyond the Proteobacteria. In *Biochimica et biophysica acta* 1853 (6), pp. 1464–1469.

Pappas, G.; Kiriaze, I. J.; Falagas, M. E. (2008): Insights into infectious disease in the era of Hippocrates. In *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases* 12 (4), pp. 347–350.

Parfrey, L. W.; Grant, J.; Tekle, Y. I.; Lasek-Nesselquist, E.; Morrison, H. G.; Sogin, M. L. et al. (2010): Broadly sampled multigene analyses yield a well-resolved eukaryotic tree of life. In *Systematic biology* 59 (5), pp. 518–533.

Park, J. S.; Myung, G.; Kim, H. S.; Cho, B. C.; Yih, W. (2007): Growth responses of the marine photosynthetic ciliate Myrionecta rubra to different cryptomonad strains. In *Aquat. Microb. Ecol.* 48, pp. 83–90.

Patron, N. J.; Inagaki, Y.; Keeling, P. J. (2007): Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages. In *Current biology : CB* 17 (10), pp. 887–891.

Patron, N. J.; Waller, R. F.; Archibald, J. M.; Keeling, P. J. (2005): Complex protein targeting to dinoflagellate plastids. In *Journal of molecular biology* 348 (4), pp. 1015–1024.

Patron, N. J.; Rogers, M. B.; Keeling, P. J. (2004): Gene replacement of fructose-1,6-bisphosphate aldolase supports the hypothesis of a single photosynthetic ancestor of chromalveolates. In *Eukaryotic cell* 3 (5), pp. 1169–1175.

Peeters, N.; Small, I. (2001): Dual targeting to mitochondria and chloroplasts. In *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1541 (1–2), pp. 54–63.

Perkins, F. O. (1976): Zoospores of the oyster pathogen, Dermocystidium marinum. 1. Fine structure of the conoid and other sporozoan-like organelles. In *Journal of Parasitology* 62 (6), pp. 959–974.

Petersen, J.; Ludewig, A.; Michael, V.; Bunk, B.; Jarek, M.; Baurain, D.; Brinkmann, H. (2014): Chromera velia, endosymbioses and the rhodoplex hypothesis--plastid evolution in cryptophytes, alveolates, stramenopiles, and haptophytes (CASH lineages). In *Genome biology and evolution* 6 (3), pp. 666–684.

Petersen, J.; Teich, R.; Brinkmann, H.; Cerff, R. (2006): A "green" phosphoribulokinase in complex algae with red plastids: evidence for a single secondary endosymbiosis leading to haptophytes, cryptophytes, heterokonts, and dinoflagellates. In *Journal of molecular evolution* 62 (2), pp. 143–157.

Philippe, H. (1993): MUST, a computer package of Management Utilities for Sequences and Trees. In *Nucleic acids research* 21 (22), pp. 5264–5272.

Philippi, M. L.; Parish, R. W.; Hohl, H. R. (1975): Histochemical and biochemical evidence for the presence of microbodies in Phytophthora palmivora. In *Arch. Microbiol.* 103 (1), pp. 127–132.

Platta, Harald W.; Hagen, Stefanie; Erdmann, Ralf (2013): The exportomer: the peroxisomal receptor export machinery. In *Cellular and molecular life sciences : CMLS* 70 (8), pp. 1393–1411.

Poinar, G. O. (2016): What Fossils Reveal About the Protozoa Progenitors, Geographic Provinces, and Early Hosts of Malarial Organisms. In *American Entomologist* 62 (1), pp. 22–25.

Poinar, G. O. (2012): Fossil gregarines in Dominican and Burmese amber: examples of accelerated development? In *Palaeodiversity* 5, pp. 1–6.

Poinar, G.; Telford, S. R. (2005): Paleohaemoproteus burmacis gen. n., sp. n. (Haemospororida. Plasmodiidae) from an Early Cretaceous biting midge (Diptera: Ceratopogonidae). In *Parasitology* 131 (1), pp. 79–84.

Poinar, G. O. (2005): Culex malariager, n. sp.(Diptera: Culicidae) from Dominican amber: The first fossil mosquito vector of Plasmodium. In *Proceedings of the Entomological Society of Washington* 107 (3), pp. 548–553.

Poinar, G. O. (2005): Plasmodium dominicana n. sp.(Plasmodiidae: Haemospororida) from Tertiary Dominican amber. In *Systematic parasitology* 61 (1), pp. 47–52.

Poinar, G. O.; Waggoner, B. M.; Bauer, U.-C. (1993): Terrestrial soft-bodied protists and other microorganisms in Triassic amber. In *Science* 259 (5092), pp. 222–224.

Polonais, V.; Soldati-Favre, D. (2010): Versatility in the acquisition of energy and carbon sources by the Apicomplexa. In *Biology of the cell / under the auspices of the European Cell Biology Organization* 102 (8), pp. 435–445.

Preiser, P. R.; Wilson, R. J.; Moore, P. W.; McCready, S.; Hajibagheri, M. A.; Blight, K. J. et al. (1996): Recombination associated with replication of malarial mitochondrial DNA. In *The EMBO journal* 15 (3), pp. 684–693.

Prescott, D. M. (1994): The DNA of ciliated protozoa. In *Microbiological reviews* 58 (2), pp. 233–267.

Pritchard, A. E.; Seilhamer, J. J.; Mahalingam, R.; Sable, C. L.; Venuti, S. E.; Cummings, D. J. (1990): Nucleotide sequence of the mitochondrial genome of Paramecium. In *Nucleic acids research* 18 (1), pp. 173–180.

Qi, W.; Li, J.; Chain, C. Y.; Pasquevich, G. A.; Pasquevich, A. F.; Cowan, J. A. (2012): Glutathione complexed Fe-S centers. In *Journal of the American Chemical Society* 134 (26), pp. 10745–10748.

Rachubinski, R. A.; Subramani, S. (1995): How proteins penetrate peroxisomes. In *Cell* 83 (4), pp. 525–528.

Ralph, S. A.; van Dooren, G. G.; Waller, R. F.; Crawford, M. J.; Fraunholz, M. J.; Foth, B. J. et al. (2004): Tropical infectious diseases: metabolic maps and functions of the Plasmodium falciparum apicoplast. In *Nature reviews. Microbiology* 2 (3), pp. 203–216.

Ramakrishnan, S.; Docampo, M. D.; MacRae, J. I.; Pujol, F. M.; Brooks, C. F.; van Dooren, G. G. et al. (2012): Apicoplast and endoplasmic reticulum cooperate in fatty acid biosynthesis in apicomplexan parasite Toxoplasma gondii. In *The Journal of biological chemistry* 287 (7), pp. 4957–4971.

Reddy, J. K.; Mannaerts, G. P. (1994): Peroxisomal lipid metabolism. In *Annual review of nutrition* 14 (1), pp. 343–370.

Reyes-Prieto, A.; Moustafa, A.; Bhattacharya, D. (2008): Multiple genes of apparent algal origin suggest ciliates may once have been photosynthetic. In *Current biology : CB* 18 (13), pp. 956–962.

Reyes-Prieto, A.; Weber, A. P. M.; Bhattacharya, D. (2007): The origin and establishment of the plastid in algae and plants. In *Annual review of genetics* 41, pp. 147–168.

Rhodin, J. (1954): Correlation of ultrastructural organization and function in normal experimentally changed convoluted tubule cells of the mouse kidney. In *Aktiebologet Godvil, Sweden*.

Ricard, G.; McEwan, N. R.; Dutilh, B. E.; Jouany, J.-P.; Macheboeuf, D.; Mitsumori, M. et al. (2006): Horizontal gene transfer from Bacteria to rumen Ciliates indicates adaptation to their anaerobic, carbohydrates-rich environment. In *BMC genomics* 7, p. 22.

Rice, P.; Longden, I.; Bleasby, A. (2000): EMBOSS: the European Molecular Biology Open Software Suite. In *Trends in genetics : TIG* 16 (6), pp. 276–277.

Riordan, C. E.; Langreth, S. G.; Sanchez, L. B.; Kayser, O.; Keithly, J. S. (1999): Preliminary evidence for a mitochondrion in Cryptosporidium parvum: phylogenetic and therapeutic implications. In *The Journal of eukaryotic microbiology* 46 (5), 52S-55S.

Roberts, C. W.; Roberts, F.; Henriquez, F. L.; Akiyoshi, D.; Samuel, B. U.; Richards, T. A. et al. (2004): Evidence for mitochondrial-derived alternative oxidase in the apicomplexan parasite Cryptosporidium parvum: a potential anti-microbial agent target. In *International journal for parasitology* 34 (3), pp. 297–308.

Rodriguez-Ezpeleta, N.; Brinkmann, H.; Burey, S. C.; Roure, B.; Burger, G.; Löffelhardt, W. et al. (2005): Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. In *Current biology : CB* 15 (14), pp. 1325–1330.

Rogers, M. B.; Gilson, P. R.; Su, V.; McFadden, G. I.; Keeling, P. J. (2007): The complete chloroplast genome of the chlorarachniophyte Bigelowiella natans: evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts. In *Molecular biology and evolution* 24 (1), pp. 54–62.

Rogozin, I. B.; Carmel, L.; Csuros, M.; Koonin, E. V. (2012): Origin and evolution of spliceosomal introns. In *Biology direct* 7, p. 11. DOI: 10.1186/1745-6150-7-11.

Rohdich, F.; Hecht, S.; Gartner, K.; Adam, P.; Krieger, C.; Amslinger, S. et al. (2002): Studies on the nonmevalonate terpene biosynthetic pathway: metabolic role of IspH (LytB) protein. In *Proceedings of the National Academy of Sciences of the United States of America* 99 (3), pp. 1158–1163.

Rohmer, M.; Knani, M.; Simonin, P.; Sutter, B.; Sahm, H. (1993): Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate. In *The Biochemical journal* 295 (Pt 2), pp. 517–524.

Roure, B.; Rodriguez-Ezpeleta, N.; Philippe, H. (2007): SCaFoS: a tool for selection, concatenation and fusion of sequences for phylogenomics. In *BMC evolutionary biology* 7 Suppl 1, S2.

Rucktäschel, R.; Girzalsky, W.; Erdmann, R. (2011): Protein import machineries of peroxisomes. In *Biochimica et biophysica acta* 1808 (3), pp. 892–900.

Saccone, C.; Gissi, C.; Lanave, C.; Larizza, A.; Pesole, G.; Reyes, A. (2000): Evolution of the mitochondrial genetic system: an overview. In *Gene* 261 (1), pp. 153–159.

Saldarriaga, J. F.; McEwan, M. L.; Fast, N. M.; Taylor, F. J. R.; Keeling, P. J. (2003): Multiple protein phylogenies show that Oxyrrhis marina and Perkinsus marinus are early branches of the dinoflagellate lineage. In *International journal of systematic and evolutionary microbiology* 53 (Pt 1), pp. 355–365.

Saldarriaga, J. F.; Taylor, F. J.; Keeling, P. J.; Cavalier-Smith, T. (2001): Dinoflagellate nuclear SSU rRNA phylogeny suggests multiple plastid losses and replacements. In *Journal of molecular evolution* 53 (3), pp. 204–213.

Sanchez Puerta, M. V.; Lippmeier, J. C.; Apt, K. E.; Delwiche, C. F. (2007): Plastid genes in a non-photosynthetic dinoflagellate. In *Protist* 158 (1), pp. 105–117.

Sanchez Puerta, M. V.; Bachvaroff, T. R.; Delwiche, C. F. (2005): The complete plastid genome sequence of the haptophyte Emiliania huxleyi: a comparison to other plastid genomes. In *DNA research : an international journal for rapid publication of reports on genes and genomes* 12 (2), pp. 151–156.

Sarjeant, W. A. S. (1978): Arpylorus antiquus Calandra, emend., a dinoflagellate cyst from the Upper Silurian. In *Palynology* 2 (1), pp. 167–179.

Sato, S.; Nakamura, Y.; Kaneko, T.; Asamizu, E.; Tabata, S. (1999): Complete structure of the chloroplast genome of Arabidopsis thaliana. In *DNA Research* 6 (5), pp. 283–290.

Scheibe, R. (2004): Malate valves to balance cellular energy supply. In *Physiologia plantarum* 120 (1), pp. 21–26.

Schimper, A. F. W. (1883): Ueber die entwickelung der chlorophyllkoerner und farbkoerper. In *Botanische Zeitung* 41, pp. 105–120.

Schliebs, W.; Girzalsky, W.; Erdmann, R. (2010): Peroxisomal protein import and ERAD: variations on a common theme. In *Nature reviews. Molecular cell biology* 11 (12), pp. 885–890.

Schlüter, A.; Real-Chicharro, A.; Gabaldon, T.; Sanchez-Jimenez, F.; Pujol, A. (2010): PeroxisomeDB 2.0: an integrative view of the global peroxisomal metabolome. In *Nucleic acids research* 38 (Database issue), D800-5.

Schlüter, A.; Fourcade, S.; Ripp, R.; Mandel, J. L.; Poch, O.; Pujol, A. (2006): The evolutionary origin of peroxisomes: an ER-peroxisome connection. In *Molecular biology and evolution* 23 (4), pp. 838–845.

Schmidt, A. R.; Ragazzi, E.; Coppellotti, O.; Roghi, G. (2006): A microworld in Triassic amber. In *Nature* 444 (7121), p. 835.

Schmieder, R.; Edwards, R. (2011): Quality control and preprocessing of metagenomic datasets. In *Bioinformatics (Oxford, England)* 27 (6), pp. 863–864.

Schmieder, Robert (2011): Quality control with PRINSEQ. Available online at http://prinseq.sourceforge.net/Quality_control_with_PRINSEQ.pdf.

Schnarrenberger, C.; Martin, W. (2002): Evolution of the enzymes of the citric acid cycle and the glyoxylate cycle of higher plants. A case study of endosymbiotic gene transfer. In *European journal of biochemistry / FEBS* 269 (3), pp. 868–883.

Schnell, D. J.; Blobel, G.; Keegstra, K.; Kessler, F.; Ko, K.; Soll, J. (1997): A consensus nomenclature for the protein-import components of the chloroplast envelope. In *Trends in cell biology* 7 (8), pp. 303–304.

Schönborn, W.; Dörfelt, H.; Foissner, W.; Krienitz, L.; Schäfer, U. (1999): A Fossilized Microcenosis In Triassic Amber. In *J Eukaryotic Microbiology* 46 (6), pp. 571–584.

Schrader, M.; Fahimi, H. D. (2008): The peroxisome: still a mysterious organelle. In *Histochemistry and cell biology* 129 (4), pp. 421–440.

Seaman, G. R. (1970): Activity of the glyoxylate cycle in Tetrahymena after infection of the cockroach, Periplanetal. In *Experimental Parasitology* 27 (1), pp. 15–21.

Seeber, F.; Soldati-Favre, D. (2010): Metabolic Pathways in the Apicoplast of Apicomplexa. In, vol. 281: Elsevier (International Review of Cell and Molecular Biology), pp. 161–228.

Seeber, F.; Limenitakis, J.; Soldati-Favre, D. (2008): Apicomplexan mitochondrial metabolism: a story of gains, losses and retentions. In *Trends in parasitology* 24 (10), pp. 468–478.

Shen, Y. Q.; Burger, G. (2009): Plasticity of a key metabolic pathway in fungi. In *Functional & integrative genomics* 9 (2), pp. 145–151.

Shiflett, A. M.; Johnson, P. J. (2010): Mitochondrion-related organelles in eukaryotic protists. In *Annual review of microbiology* 64, pp. 409–429.

Shoguchi, E.; Shinzato, C.; Hisata, K.; Satoh, N.; Mungpakdee, S. (2015): The Large Mitochondrial Genome of Symbiodinium minutum Reveals Conserved Noncoding Sequences between Dinoflagellates and Apicomplexans. In *Genome biology and evolution* 7 (8), pp. 2237–2244.

Shoguchi, E.; Shinzato, C.; Kawashima, T.; Gyoja, F.; Mungpakdee, S.; Koyanagi, R. et al. (2013): Draft assembly of the Symbiodinium minutum nuclear genome reveals dinoflagellate gene structure. In *Current biology : CB* 23 (15), pp. 1399–1408.

Siddall, M. E.; Reece, K. S.; Graves, J. E.; Burreson, E. M. (1997): 'Total evidence' refutes the inclusion of Perkinsus species in the phylum Apicomplexa. In *Parasitology* 115 (2), pp. 165–176.

Siekevitz, P. (1957): Powerhouse of the Cell. In *Scientific American* 197, pp. 131–144.

Sims, P. A.; Mann, D. G.; Medlin, L. K. (2006): Evolution of the diatoms. Insights from fossil, biological and molecular data. In *Phycologia* 45 (4), pp. 361–402.

Siregar, J. E.; Kurisu, G.; Kobayashi, T.; Matsuzaki, M.; Sakamoto, K.; Mi-ichi, F. et al. (2015): Direct evidence for the atovaquone action on the Plasmodium cytochrome bc1 complex. In *Parasitology international* 64 (3), pp. 295–300.

Slamovits, C. H.; Keeling, P. J. (2008): Plastid-derived genes in the nonphotosynthetic alveolate Oxyrrhis marina. In *Molecular biology and evolution* 25 (7), pp. 1297–1306.

Slamovits, C. H.; Saldarriaga, J. F.; Larocque, A.; Keeling, P. J. (2007): The highly reduced and fragmented mitochondrial genome of the early-branching dinoflagellate Oxyrrhis marina shares characteristics with both apicomplexan and dinoflagellate mitochondrial genomes. In *Journal of molecular biology* 372 (2), pp. 356–368.

Smith, D. R.; Lee, R. W. (2014): A plastid without a genome: evidence from the nonphotosynthetic green algal genus Polytomella. In *Plant physiology* 164 (4), pp. 1812–1819.

Smith, J. J.; Aitchison, J. D. (2013): Peroxisomes take shape. In *Nature reviews. Molecular cell biology* 14 (12), pp. 803–817.

Smith, J. J.; Aitchison, J. D. (2009): Regulation of peroxisome dynamics. In *Current opinion in cell biology* 21 (1), pp. 119–126.

Sonnenborn, U.; Kunau, W.-H. (1982): Purification and properties of the fatty acid synthetase complex from the marine dinoflagellate, Crypthecodinium cohnii. In *Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism* 712 (3), pp. 523–534.

Souza, W. de; Attias, M.; Rodrigues, J. C. F. (2009): Particularities of mitochondrial structure in parasitic protists (Apicomplexa and Kinetoplastida). In *The international journal of biochemistry & cell biology* 41 (10), pp. 2069–2080.

Speijer, D. (2015): Birth of the eukaryotes by a set of reactive innovations: New insights force us to relinquish gradual models. In *BioEssays : news and reviews in molecular, cellular and developmental biology* 37 (12), pp. 1268–1276.

Srivastava, I. K.; Rottenberg, H.; Vaidya, A. B. (1997): Atovaquone, a Broad Spectrum Antiparasitic Drug, Collapses Mitochondrial Membrane Potential in a Malarial Parasite. In *Journal of Biological Chemistry* 272 (7), pp. 3961–3966.

Stamatakis, A.; Alachiotis, N. (2010): Time and memory efficient likelihood-based tree searches on phylogenomic alignments with missing data. In *Bioinformatics (Oxford, England)* 26 (12), i132-9.

Stamati, K.; Mudera, V.; Cheema, U. (2011): Evolution of oxygen utilization in multicellular organisms and implications for cell signalling in tissue engineering. In *Journal of tissue engineering* 2 (1), 2041731411432365.

Stelter, K.; El-Sayed, N. M.; Seeber, F. (2007): The expression of a plant-type ferredoxin redox system provides molecular evidence for a plastid in the early dinoflagellate Perkinsus marinus. In *Protist* 158 (1), pp. 119–130.

Stirewalt, V. L.; Michalowski, C. B.; Löffelhardt, W.; Bohnert, H. J.; Bryant, D. A. (1995): Nucleotide sequence of the cyanelle genome fromCyanophora paradoxa. In *Plant Mol Biol Rep* 13 (4), pp. 327–332.

Summerer, M.; Sonntag, B.; Sommaruga, R. (2008): Ciliate-symbiont specificity of freshwater endosymbiotic Chlorella (Trebouxiophyceae, Chlorophyta). In *Journal of Phycology* 44 (1), pp. 77–84.

Sunila, I.; Hamilton, R. M.; Dungan, C. F. (2001): Ultrastructural characteristics of the in vitro cell cycle of the protozoan pathogen of oysters, Perkinsus marinus. In *The Journal of eukaryotic microbiology* 48 (3), pp. 348–361.

Swart, E. C.; Nowacki, M.; Shum, J.; Stiles, H.; Higgins, B. P.; Doak, T. G. et al. (2012): The Oxytricha trifallax mitochondrial genome. In *Genome biology and evolution* 4 (2), pp. 136–154.

Tabak, H. F.; Hoepfner, D.; Zand, Av; Geuze, H. J.; Braakman, I.; Huynen, M. A. (2006): Formation of peroxisomes: present and past. In *Biochimica et biophysica acta* 1763 (12), pp. 1647–1654.

Takishita, K.; Kolke, K.; Maruyama, T.; Ogata, T. (2002): Molecular evidence for plastid robbery (kleptoplastidy) in Dinophysis, a dinoflagellate causing diarrhetic shellfish poisoning. In *Protist* 153 (3), pp. 293–302.

Talavera, G.; Castresana, J. (2007): Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. In *Systematic biology* 56 (4), pp. 564–577.

Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. (2011): MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. In *Molecular biology and evolution* 28 (10), pp. 2731–2739.

Tanaka, R.; Tanaka, A. (2007): Tetrapyrrole biosynthesis in higher plants. In *Annual review of plant biology* 58, pp. 321–346.

Teles-Grilo, M. L.; Tato-Costa, J.; Duarte, S. M.; Maia, A.; Casal, G.; Azevedo, C. (2007): Is there a plastid in Perkinsus atlanticus (Phylum Perkinsozoa)? In *European Journal of Protistology* 43 (2), pp. 163–167.

Tengs, T.; Dahlberg, O. J.; Shalchian-Tabrizi, K.; Klaveness, D.; Rudi, K.; Delwiche, C. F.; Jakobsen, K. S. (2000): Phylogenetic analyses indicate that the 19′ hexanoyloxy-fucoxanthin-containing dinoflagellates have tertiary plastids of haptophyte origin. In *Molecular biology and evolution* 17 (5), pp. 718–729.

Thompson, J. D.; Gibson, T. J.; Plewniak, F.; Jeanmougin, F.; Higgins, D. G. (1997): The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. In *Nucleic acids research* 25 (24), pp. 4876–4882.

Tikhonenkov, D. V.; Janouskovec, J.; Mylnikov, A. P.; Mikhailov, K. V.; Simdyanov, T. G.; Aleoshin, V. V.; Keeling, P. J. (2014): Description of Colponema vietnamica sp.n. and Acavomonas peruviana n. gen. n. sp., two new alveolate phyla (Colponemidia nom. nov. and Acavomonidia nom. nov.) and their contributions to reconstructing the ancestral state of alveolates and eukaryotes. In *PloS one* 9 (4), e95467.

Timmis, J. N.; Ayliffe, M. A.; Huang, C. Y.; Martin, W. (2004): Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. In *Nature reviews. Genetics* 5 (2), pp. 123–135.

Tolbert, N. E.; Essner, E. (1981): Microbodies: peroxisomes and glyoxysomes. In *The Journal of Cell Biology* 91 (3 Pt 2), 271s-283s.

Toso, M. A.; Omoto, C. K. (2007): Gregarina niphandrodes may lack both a plastid genome and organelle. In *The Journal of eukaryotic microbiology* 54 (1), pp. 66–72.

Tovar, J.; Leon-Avila, G.; Sanchez, L. B.; Sutak, R.; Tachezy, J.; van der Giezen, M. et al. (2003): Mitochondrial remnant organelles of Giardia function in iron-sulphur protein maturation. In *Nature* 426 (6963), pp. 172–176.

Tovar, J.; Fischer, A.; Clark, C. G. (1999): The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite Entamoeba histolytica. In *Mol Microbiol* 32 (5), pp. 1013–1021.

Trumpower, B. L. (1990): The protonmotive Q cycle. Energy transduction by coupling of proton translocation to electron transfer by the cytochrome bc1 complex. In *The Journal of biological chemistry* 265 (20), pp. 11409–11412.

Tsaousis, A. D.; Gentekaki, E.; Eme, L.; Gaston, D.; Roger, A. J. (2014): Evolution of the cytosolic iron-sulfur cluster assembly machinery in Blastocystis species and other microbial eukaryotes. In *Eukaryotic cell* 13 (1), pp. 143–153.

Tsukihara, T.; Aoyama, H.; Yamashita, E.; Tomizaki, T.; Yamaguchi, H.; Shinzawa-Itoh, K. et al. (1996): The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 A. In *Science (New York, N.Y.)* 272 (5265), pp. 1136–1144.

Tu, Y. (2011): The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine. In *Nature medicine* 17 (10), pp. 1217–1220.

Turmel, M.; Gagnon, M.-C.; O'Kelly, C. J.; Otis, C.; Lemieux, C. (2009): The chloroplast genomes of the green algae Pyramimonas, Monomastix, and Pycnococcus shed new light on the evolutionary history of prasinophytes and the origin of the secondary chloroplasts of euglenids. In *Molecular biology and evolution* 26 (3), pp. 631–648.

Ucko, M.; Elbträchter, M.; Schnepf, E. (1997): A Crypthecodinium cohnii-like dinoflagellate feeding myzocytotically on the unicellular red alga Porphyridium sp. In *European Journal of Phycology* 32 (02), pp. 133–140.

van der Giezen, M. (2011): Mitochondria and the Rise of Eukaryotes. In *BioScience* 61 (8), pp. 594–601.

van der Giezen, M. (2009): Hydrogenosomes and mitosomes: conservation and evolution of functions. In *The Journal of eukaryotic microbiology* 56 (3), pp. 221–231.

van der Leij, I.; van den Berg, M.; Boot, R.; Franse, M.; Distel, B.; Tabak, H. F. (1992): Isolation of peroxisome assembly mutants from Saccharomyces cerevisiae with different morphologies using a novel positive selection procedure. In *The Journal of Cell Biology* 119 (1), pp. 153–162.

van Dolah, F. M.; Zippay, M. L.; Pezzolesi, L.; Rein, K. S.; Johnson, J. G.; Morey, J. S. et al. (2013): Subcellular localization of dinoflagellate polyketide synthases and fatty acid synthase activity. In *Journal of Phycology* 49 (6), pp. 1118–1127.

van Dooren, G. G.; Kennedy, A. T.; McFadden, G. I. (2012): The use and abuse of heme in apicomplexan parasites. In *Antioxidants & redox signaling* 17 (4), pp. 634–656.

van Dooren, G. G.; Tomova, C.; Agrawal, S.; Humbel, B. M.; Striepen, B. (2008): Toxoplasma gondii Tic20 is essential for apicoplast protein import. In *Proceedings of the National Academy of Sciences of the United States of America* 105 (36), pp. 13574–13579.

van Dooren, G. G.; Stimmler, L. M.; McFadden, G. I. (2006): Metabolic maps and functions of the Plasmodium mitochondrion. In *FEMS microbiology reviews* 30 (4), pp. 596–630.

Vaughan, A. M.; O'Neill, M. T.; Tarun, A. S.; Camargo, N.; Phuong, T. M.; Aly, A. S. I. et al. (2009): Type II fatty acid synthesis is essential only for malaria parasite late liver stage development. In *Cellular microbiology* 11 (3), pp. 506–520.

Veenhuis, M.; van der Klei, I. J. (2014): A critical reflection on the principles of peroxisome formation in yeast. In *Frontiers in physiology* 5, p. 110.

Vries, S. de; Grivell, L. A. (1988): Purification and characterization of a rotenone-insensitive NADH:Q6 oxidoreductase from mitochondria of Saccharomyces cerevisiae. In *European journal of biochemistry / FEBS* 176 (2), pp. 377–384.

Walker, J. M. (Ed.) (2005): The Proteomics Protocols Handbook. Totowa, NJ: Humana Press.

Wallace, D. C. (2005): A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. In *Annual review of genetics* 39, pp. 359–407.

Waller, R. F.; Jackson, C. J. (2009): Dinoflagellate mitochondrial genomes: stretching the rules of molecular biology. In *BioEssays : news and reviews in molecular, cellular and developmental biology* 31 (2), pp. 237–245.

Waller, R. F.; Keeling, P. J. (2006): Alveolate and chlorophycean mitochondrial cox2 genes split twice independently. In *Gene* 383, pp. 33–37.

Waller, R. F.; Keeling, P. J.; van Dooren, G. G.; McFadden, G. I. (2003): Comment on "A green algal apicoplast ancestor". In *Science (New York, N.Y.)* 301 (5629), 49; author reply 49.

Waller, R. F.; Reed, M. B.; Cowman, A. F.; McFadden, G. I. (2000): Protein trafficking to the plastid of Plasmodium falciparum is via the secretory pathway. In *The EMBO journal* 19 (8), pp. 1794–1802.

Walter, M. R.; Buick, R.; Dunlop, J. S.R. (1980): Stromatolites 3,400–3,500 Myr old from the North Pole area, Western Australia. In *Nature*, pp. 443–445.

Wanders, R. J.A.; Waterham, H. R.; Ferdinandusse, S. (2015): Metabolic Interplay between Peroxisomes and Other Subcellular Organelles Including Mitochondria and the Endoplasmic Reticulum. In *Frontiers in Cell and Developmental Biology* 3, p. 83.

Wanders, R. J. A.; Ferdinandusse, S.; Brites, P.; Kemp, S. (2010): Peroxisomes, lipid metabolism and lipotoxicity. In *Biochimica et biophysica acta* 1801 (3), pp. 272–280.

Wang, Q.; Garrity, G. M.; Tiedje, J. M.; Cole, JR (2007): Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. In *Applied and environmental microbiology* 73 (16), pp. 5261–5267.

Wang, Y.; Morse, D. (2006): Rampant polyuridylylation of plastid gene transcripts in the dinoflagellate Lingulodinium. In *Nucleic acids research* 34 (2), pp. 613–619.

Whiteman, H. (2016): Malaria dates back to dinosaur era, study suggests. Edited by Medical News Today. Available online at http://www.medicalnewstoday.com/articles/308400.php.

WHO (2016): World malaria report 2015. [S.l.]: World Health Organization.

Wilby, K. J.; Lau, T. T.; Gilchrist, S. E.; Ensom, M. H. (2012): Mosquirix (RTS,S). A Novel Vaccine for the Prevention of Plasmodium falciparum Malaria. In *Annals of Pharmacotherapy* 46 (3), pp. 384–393.

Wilcox, L. W.; Wedemayer, G. J. (1985): Dinoflagellate with blue-green chloroplasts derived from an endosymbiotic eukaryote. In *Science* 227 (4683), pp. 192–194.

Wilde, S. A.; Valley, J. W.; Peck, W. H.; Graham, C. M. (2001): Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago. In *Nature* 409 (6817), pp. 175–178.

Wilkinson, L. (2012): Exact and approximate area-proportional circular Venn and Euler diagrams. In *IEEE transactions on visualization and computer graphics* 18 (2), pp. 321–331.

Williams, B.; Keeling, P. J. (2003): Cryptic Organelles in Parasitic Protists and Fungi. In : Advances in Parasitology Volume 54, vol. 54: Elsevier (Advances in Parasitology), pp. 9–68.

Williams, C.; Distel, B. (2006): Pex13p: docking or cargo handling protein? In *Biochimica et biophysica acta* 1763 (12), pp. 1585–1591.

Wilson, P. B. M. de; Allen, J. F.; van der Giezen, M. (2012): Mitochondria, hydrogenosomes and mitosomes in relation to the CoRR hypothesis for genome function and evolution. In E. C. Bullerwell (Ed.): Organelle Genetics: Evolution of Organelle Genomes and Gene Expression. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 105–119.

Wilson, R. J. M.; Denny, P. W.; Preiser, P. R.; Rangachari, K.; Roberts, K.; Roy, A. et al. (1996): Complete Gene Map of the Plastid-like DNA of the Malaria ParasitePlasmodium falciparum. In *Journal of molecular biology* 261 (2), pp. 155–172.

Wisecaver, J. H.; Brosnahan, M. L.; Hackett, J. D. (2013): Horizontal gene transfer is a significant driver of gene innovation in dinoflagellates. In *Genome biology and evolution* 5 (12), pp. 2368–2381.

Wisecaver, J. H.; Hackett, J. D. (2011): Dinoflagellate genome evolution. In *Annual review of microbiology* 65, pp. 369–387.

Wisecaver, J. H.; Hackett, J. D. (2010): Transcriptome analysis reveals nuclear-encoded proteins for the maintenance of temporary plastids in the dinoflagellate Dinophysis acuminata. In *BMC genomics* 11, p. 366.

Woehle, C.; Dagan, T.; Martin, W. F.; Gould, S. B. (2011): Red and problematic green phylogenetic signals among thousands of nuclear genes from the photosynthetic and apicomplexa-related Chromera velia. In *Genome biology and evolution* 3, pp. 1220–1230.

Wolf, Y. I.; Koonin, E. V. (2013): Genome reduction as the dominant mode of evolution. In *Bioessays* 35 (9), pp. 829–837.

Woo, Y. H.; Ansari, H.; Otto, T. D.; Klinger, C. M.; Kolisko, M.; Michalek, J. et al. (2015): Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites. In *eLife* 4, e06974.

Xi, Z.; Wang, Y.; Bradley, R. K.; Sugumaran, M.; Marx, C. J.; Rest, J. S.; Davis, C. C. (2013): Massive mitochondrial gene transfer in a parasitic flowering plant clade. In *PLoS genetics* 9 (2), e1003265.

Xia, S.; Zhang, Q.; Zhu, H.; Cheng, Y.; Liu, G.; Hu, Z. (2013): Systematics of a kleptoplastidal dinoflagellate, Gymnodinium eucyaneum Hu (Dinophyceae), and its cryptomonad endosymbiont. In *PloS one* 8 (1), e53820.

Xu, P.; Widmer, G.; Wang, Y.; Ozaki, L. S.; Alves, J. M.; Serrano, M. G. et al. (2004): The genome of Cryptosporidium hominis. In *Nature* 431 (7012), pp. 1107–1112.

Yandell, Mark; Ence, Daniel (2012): A beginner's guide to eukaryotic genome annotation. In *Nature Reviews Genetics* 13 (5), pp. 329–342.

Yang, Y.; Matsuzaki, M.; Takahashi, F.; Qu, L.; Nozaki, H. (2014): Phylogenomic analysis of "red" genes from two divergent species of the "green" secondary phototrophs, the chlorarachniophytes, suggests multiple horizontal gene transfers from the red lineage before the divergence of extant chlorarachniophytes. In *PloS one* 9 (6), e101158.

Yeh, E.; DeRisi, J. L. (2011): Chemical rescue of malaria parasites lacking an apicoplast defines organelle function in blood-stage Plasmodium falciparum. In *PLoS biology* 9 (8), e1001138.

Yih, W.; Kim, H. S.; Jeong, H. J.; Myung, G.; Kim, Y. G. (2004): Ingestion of cryptophyte cells by the marine photosynthetic ciliate Mesodinium rubrum. In *Aquat. Microb. Ecol.* 36, pp. 165–170.

Yoon, H. S.; Hackett, J. D.; van Dolah, F. M.; Nosenko, T.; Lidie, K. L.; Bhattacharya, D. (2005): Tertiary endosymbiosis driven genome evolution in dinoflagellate algae. In *Molecular biology and evolution* 22 (5), pp. 1299–1308.

Yoon, H. S.; Hackett, J. D.; Ciniglia, C.; Pinto, G.; Bhattacharya, D. (2004): A molecular timeline for the origin of photosynthetic eukaryotes. In *Molecular biology and evolution* 21 (5), pp. 809–818..

Yoon, T.; Cowan, J. A. (2003): Iron-sulfur cluster biosynthesis. Characterization of frataxin as an iron donor for assembly of 2Fe-2S clusters in ISU-type proteins. In *Journal of the American Chemical Society* 125 (20), pp. 6078–6084.

Yoon, H. S.; Hackett, J. D.; Pinto, G.; Bhattacharya, D. (2002): The Single, Ancient Origin of Chromist Plastids. In *J Phycol* 99 (24), pp. 15507–15512.

Yotoko, K. S. C.; Elisei, C. (2006): Malaria parasites (Apicomplexa, Haematozoea) and their relationships with their hosts. Is there an evolutionary cost for the specialization? In *Journal of Zoological Systematics and Evolutionary Research* 44 (4), pp. 265–273.

Zauner, S.; Greilinger, D.; Laatsch, T.; Kowallik, K. V.; Maier, U.-G. (2004): Substitutional editing of transcripts from genes of cyanobacterial origin in the dinoflagellate Ceratium horridum. In *FEBS letters* 577 (3), pp. 535–538.

Zeidler, J.; Schwender, J.; Mueller, C.; Lichtenthaler, H. K. (2000): The non-mevalonate isoprenoid biosynthesis of plants as a test system for drugs against malaria and pathogenic bacteria. In *Biochem. Soc. Trans* 28 (6), pp. 796–798.

Zerbino, D. R.; Birney, E. (2008): Velvet: algorithms for de novo short read assembly using de Bruijn graphs. In *Genome research* 18 (5), pp. 821–829.

Zhang, H.; Campbell, D. A.; Sturm, N. R.; Dungan, C. F.; Lin, S. (2011): Spliced leader RNAs, mitochondrial gene frameshifts and multi-protein phylogeny expand support for the genus Perkinsus as a unique group of alveolates. In *PloS one* 6 (5), e19933.

Zhang, H.; Hou, Y.; Miranda, L.; Campbell, D. A.; Sturm, N. R.; Gaasterland, T.; Lin, S. (2007): Spliced leader RNA trans-splicing in dinoflagellates. In *Proceedings of the National Academy of Sciences of the United States of America* 104 (11), pp. 4618–4623..

Zhang, Z.; Cavalier-Smith, T.; Green, B. R. (2002): Evolution of dinoflagellate unigenic minicircles and the partially concerted divergence of their putative replicon origins. In *Molecular biology and evolution* 19 (4), pp. 489–500.

Zhang, Z.; Green, B. R.; Cavalier-Smith, T. (1999): Single gene circles in dinoflagellate chloroplast genomes. In *Nature* 400 (6740), pp. 155–159.

Zhu, G. (2004): Current progress in the fatty acid metabolism in Cryptosporidium parvum. In *The Journal of eukaryotic microbiology* 51 (4), pp. 381–388.

Zhu, G.; Marchewka, M. J.; Keithly, J. S. (2000): Cryptosporidium parvum appears to lack a plastid genome. In *Microbiology* 146 (2), pp. 315–321.

List of Abbreviations

aa	amino acid
ABCD	ATP-binding cassette, subfamily D, member 1
ACAA1	acetyl-CoA acyltransferase 1
ACO	aconitase
ACOX	acetyl-CoA oxidase
ACSL	long-chain acyl-CoA synthetase
AGPS	alkyldihydroxyacetone phosphate synthase
ALA	5-aminolevulinic acid
ALAS	5-aminolevulinic acid synthase
Amp	ampicillin
AOX	alternative oxidase
ARFTM	Alternative Ray's Fluid Thioglycollate Medium
ATCC	American Type Culture Collection
ATP	adenosine triphosphate
BC	before Christ
BLAST	Basic Local Alignment Search Tool
bp	base pair
BS	bootstrap support
BSA	bovine serum albumin
°C	degree Celsius
Cam	chloramphenicol
CASH	Cryptophytes, Alveolates, Stramenopiles, Haptophytes
CAT	catalase
CC	cellular component
CCAP	Scottish Culture Collection of Algae and Protozoa
CCMP	Provasoli-Guillard National Center for Culture of Marine Phytoplankton
CMK	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase
cDNA	complementary DNA
CoA	coenzyme A
cob	cytochrome b
CoRR	co-location for redox-regulation
cox	cytochrome c oxidase
CS	citrate synthase
CsCl	cesium chloride
DAO	D-amino-acid oxidase
dATP	deoxyadenosine triphosphate
DBP	enoyl-CoA hydratase/3-hydroxy-acyl-CoA dehydrogenase (multifunctional protein)
dCTP	deoxycytidine triphosphate
DHAP	dihydroxyacetone phosphate
DHAPAT	dihydroxyacetone phosphate acyltransferase
DMAPP	dimethylallyl pyrophosphate
DME	Dulbecco's Modified Eagle's Medium
DNA	deoxyribonucleic acid
dNTP	desoxyribonucleosid triphosphate
dT	deoxythymidine
DXR	1-deoxy-D-xylulose 5-phosphate reductoisomerase

DXS	1-deoxy-D-xylulose-5-phosphate synthase
ECH	Delta(3,5)-Delta(2,4)-dienoyl-CoA isomerase
EDTA	ethylenediaminetetraacetic acid
e.g.	"exempli gratia", for example
EGT	endosymbiotic gene transfer
ERAD	endoplasmatic-reticulum-associated protein degradation
EST	expressed sequence tag
ETC	electron transport chain
FabG	3-oxoacyl-[acyl-carrier-protein] reductase
FAR	fatty acyl reductase
FAS	fatty acid synthesis
Fe-S	iron-sulphur
µg	microgram
Ga	billion years
GA	Genome Analyzer (Illumina)
GC	guanosine/cytosine
gDNA	genomic DNA
GO	Gene Ontology
GTR	generalised time reversible
h	hour
H2O2	hydrogen peroxide
HAO	(S)-2-hydroxy-acid oxidase
HDR	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
HDS	4-hydroxy-3-methylbut-2-en-1-yl disphosphate synthase
HemB	porphobilinogen synthase
HemC	porphobilinogen deaminase
HemD	uroporphyrinogen III synthase
HemE	uroporphyrinogen III decarboxylase
HemF	coproporphyrinogen oxidase
HemG/Y	protoporphyrinogen oxidase
HemH	ferrochelatase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HGT	horizontal gene transfer
HMB-PP	4-hydroxy-3-methylbut-2-enyl pyrophosphate
HPCL2	2-hydroxyacyl-CoA lyase
HRP	horseradish peroxidase
ICL	isocitrate lyase
IgG	immunoglobulin G
iPCR	inverse PCR
IPP	isopentenyl pyrophosphate
IPTG	isopropyl ?-D-1-thiogalactopyranoside
KAAS	KEGG automated annotation server
kb	kilobase
kDa	kilodalton
KEGG	Kyoto Encyclopedia of Genes and Genomes
kg	kilogram
KO	KEGG ortholog
µL	microliter
L	liter
LB	lysogeny broth
LG	model substitution model designed by S. Q. Le and O. Gascuel

LM	light microscopy
log phase	logarithmic/exponential phase in bacterial growth
LSU	rRNA large-subunit ribosomal RNA
µM	micromolar
MCT	2-C-methyl-D-erythritol 4-phosphate cytidiylyltransferase
MDH	malate dehydrogenase
MEcPP	2-C-methyl-D-erythritol 2,4-cyclodiphosphate
MECPS	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase
MEP	2-C-methyl-D-erythritol 4-phosphate
min	minute
mL	milliliter
MLS	malate synthase
mM	millimolar
mOsm	milliosmole
mRNA	messenger RNA
MRO	mitochondria-related organelles
MSA	multiple sequence alignment
MVA	mevalonic acid
mya	million years ago
NaCl	sodium chloride
NADH	reduced nicotinamide adenine dinucleotide
NADP+	nicotinamide adenine dinucleotide phosphate
NCBI	National Center for Biotechnology Information
NDH2	alternative NADH dehydrogenase
ng	nanogram
NGS	next generation sequencing
nm	nanometer
OD	optical density
OTU	operational taxonomic unit
OXPHOS	oxidative phosphorylation
PAGE	polyacrylamide gel electrophoresis
PAOX	N1-acetylpolyamine oxidase
PBS	phosphate-buffered saline
PBS-T	PBS buffer + Tween-20
PCR	polymerase chain reaction
Pex	peroxin
PIPOX	sarcosine oxidase/L-pipecolate oxidase
PMP	peroxisomal membrane protein
ppt	parts per trillion
PRDX	peroxiredoxin
PTS	peroxisomal targeting sequence
Q	coenzyme Q10
QH2	ubiquinol
RDP	Ribosomal Database Project
RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid
rpm	revolutions per minute
SBP	sedoheptulose-1,7-bisphosphatase
SDS	sodium dodecyl sulfate
SELMA	symbiont-specific ERAD-like machinery
SL	spliced leader

SOD	superoxide dismutase
SSU rRNA	small-subunit ribosomal RNA
Taq	Thermus aquaticus
TCA	tricarboxylic acid
TdT	terminal deoxynucleotidyl transferase
TIC	translocon on the inner chloroplast membrane
TOC	translocon on the outer chloroplast membrane
tRNA	transfer ribonucleic acid
U	unit
UV	ultraviolet
V	Volt
WAG	substitution model designed by S. Whelan and N. Goldman
XDH	xanthine dehydrogenase/oxidase
xg	x gravity (measure of centrifugal force)
X-gal	5-bromo-4-chloro-3-indolyl-?-D-galactopyranoside
ycf	hypothetical chloroplast open reading frame

List of Figures

List of Tables

Danksagung

Eine Doktorarbeit ist immer auch von den Personen abhängig, die auf dem Weg zu ihrer Vervollständigung und die zum Gelingen des Manuskripts beigetragen haben.
Aus diesem Grund möchte ich mich in erster Linie bei meinem Mentor PD Dr. Jörn Petersen bedanken, der mir das Arbeiten an dieser spannenden Thematik ermöglichte, immer eine offene Tür für konstruktive Diskussionen hatte und eine Kanne Kaffee zur Arbeitsmotivation bereithielt.
Außerdem bedanke ich mich bei Prof. Dr. Michael Steinert für die Übernahme des Zweitgutachtens, sowie Prof. Dr. Miguel Vences für den Vorsitz der Prüfungskommission. Außerdem bedanke ich mich bei Prof. Dr. Miguel Vences, dass er mich in meiner gesamten universitären Laufbahn stets unterstützt und gefördert hat.
Der Deutschen Forschungsgemeinschaft danke ich für die Finanzierung der Arbeit im Rahmen des Projekts „Von der Malaria zum Meeresleuchten – Evolution der Plastiden in Alveolata“ (PE 894/2-1). Des Weiteren bedanke ich mich für die freundliche Mitintegration in den Sonderforschungsbereich SFB TRR51 „Roseobacter“, der mich als ‚eukaryotischen Außenseiter‘ wissenschaftliche Kontakte und neue Freundschaften schließen ließ. Der Universität Oldenburg danke ich für das Stipendium in meinem ersten Promotionsjahr.
Vor allem gilt mein Dank denjenigen, die an meinem Projekt mitgewirkt haben. Dazu gehört zunächst unser Universalgenie Victoria Michael, die mich nicht nur in diversen Labormethoden und Abläufen an der DSMZ unterstützt hat, sondern auch eine gute Freundin geworden ist. Für die Durchführung der Illumina Sequenzierungen, sowie für die Verwaltung und Assemblierung unserer Hochdurchsatz-Sequenzierungsdaten bedanke ich mich recht herzlich bei der Arbeitsgruppe „Genomanalytik“ des Helmholtz-Zentrums für Infektionsforschung in Braunschweig, insbesondere bei Michael Jarek. Boyke Bunk danke ich für die Teilnahme an meinem Promotionsbeirat und für die stetige Unterstützung in bioinformatischen Fragen und Analysen. Unserem Postdoc Dr. Henner Brinkmann danke ich für die Berechnung der Phylogenien, für konstruktive Diskussionen, sowie der Betreuung meines Auslandspraktikums in Montréal. Prof. Dr. Hervé Philippe danke ich für die Durchführung der phylogenomischen Analysen, sowie für die Möglichkeit, diese Methode selbst in seiner Arbeitsgruppe in Montréal erlernt haben zu dürfen. Für die Kooperation in der Kultivierung und Sequenzierung des Dinoflagellaten *P. minimum*, sowie für die Teilnahme an meinem Promotionsbeirat bedanke ich mich bei Prof. Dr. Irene Wagner-Döbler. Bei Prof. Dr.

Manfred Rohde und Ina Schleicher bedanke ich mich für Ihre Kooperation bei den elektronenmikroskopischen Experimenten. Prof. Dr. Guan Zhu danke ich für die freundliche Überlassung der Daten zu den Gregarinen. Außerdem bedanke ich mich bei Christopher F. Dungan für die Bereitstellung des *Perkinsus*-Stamms PRA-181 für dieses Projekt.

Für das fleißige Korrekturlesen und für Anregungen zu dieser Arbeit bedanke ich mich bei meinem Vater, meinem Ehemann und meiner Schwester.

Ich möchte mich bei der gesamten Arbeitsgruppe für das nette Arbeitsklima, ihre Unterstützung, die vielen Kuchen, den vielen Kaffee und den Spaß bei der Arbeit bedanken: PD Dr. Jörn Petersen, Dr. Silke Pradella, Dr. Henner Brinkmann, Orsola Päuker, Victoria Michael, Oliver Frank, Nora Buddruhs, Claire Ellebrandt und Pascal Bartling.

Bei meinen Freunden bedanke ich mich für die nötige Ablenkung, ihre Unterstützung und dass sie immer für mich da gewesen sind. Ich hatte eine wirklich tolle und prägende Zeit in Braunschweig und werde die blau-gelbe Stadt sehr vermissen.

Die Besten kommen zum Schluss: Ich bedanke mich bei meinen Eltern, meiner Schwester und meinem Ehemann dafür, dass sie mich stets in allen Lebenslagen unterstützen und mir meinen Lebensweg so weit wie nur möglich ebnen. Ohne Eure seelische Unterstützung wäre diese Arbeit nicht möglich gewesen.